(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 March 2001 (29.03.2001)

PCT

(10) International Publication Number
WO 01/21807 A1

(51) International Patent Classification[7]: C12N 15/40, 5/10, 7/01, C07K 14/18, 16/10, A61K 31/70, 39/29

(21) International Application Number: PCT/US00/25987

(22) International Filing Date:
22 September 2000 (22.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/155,823 23 September 1999 (23.09.1999) US

(71) Applicant *(for all designated States except US)*: **THE GOVERNMENT OF THE UNITED STATES OF AMERICA as represented by THE SECRETARY, DEPARTMENT OF HEALTH SERVICES** [US/US]; National Institutes of Health, Office of Technology Transfer, Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FORNS, Xavier** [US/US]; Apartment 1106, 4400 East-West Highway, Bethesda, MD 20814 (US). **BUKH, Jens** [DK/US]; 2018 Baltimore Road, #J42, Rockville, MD 20851 (US). **EMERSON, Suzanne, U.** [US/US]; 9903 Parkwood Drive, Bethesda, MD 20814 (US). **PURCELL, Robert, H.** [US/US]; 17517 White Ground Road, Boyds, MD 20841 (US).

(74) Agent: **FEILER, William, S.**; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/21807 A1

(54) Title: HEPATITIS C VIRUS ENVELOPE TWO PROTEIN (E2) WHICH LACKS ALL OR PART OF THE HYPERVARIABLE REGION ONE (HVR1), CORRESPONDING NUCLEIC ACIDS, CHIMERIC VIRUSES AND USES THEREOF

(57) Abstract: The present invention relates to nucleic acid molecules encoding hepatitis C virus, chimeric hepatitis C virus or hepatitis C virus envelope two protein which lacks all or part of hypervariable region one of the envelope two protein. The invention further relates to the use of these nucleic acid molecules and their encoded polypeptides as vaccine candidates.

Title of Invention

HEPATITE C VIRUS ENVELOPE TWO PROTEIN (E2) WICH LACKS ALL OR PART OF THE HYPERVARIABLE REGION ONE (HVR1) , CORRESPONDING NUCLEIC ACIDS, CHIMERIC VIRUSES AND USES THEREOF

Field of Invention

The present invention relates to nucleic acid molecules that encode a hepatitis C virus (HCV) envelope two protein which lacks all or part of the hypervariable region one (HVR1) of the envelope two (E2) protein. The invention further relates to the use of the nucleic acid molecules and their encoded polypeptides as vaccine candidates.

Background of Invention

Hepatitis C virus (HCV) is a positive-sense single-strand RNA virus belonging to the *Flaviviridae* family of viruses (Rice, 1996).

At present, more than 80% of individuals infected with HCV become chronically infected and these chronically infected individuals have a relatively high risk of developing chronic hepatitis, liver cirrhosis and hepatocellular carcinoma (Hoofnagle, 1997). The only effective therapy for chronic hepatitis C, interferon (IFN), alone or in combination with ribavirin, induces a sustained response in less than 50% of treated patients (Davis et al., 1998; McHutchinson et al., 1998). Consequently, HCV is currently the most common cause of end-stage liver failure and the reason for about 30% of liver transplants performed in the U.S. (Hoofnagle, 1997). As a result of the inability to develop a universally effective therapy against HCV infection, it is estimated that there are still more than 25,000 new infections yearly in the U.S. (Alter

1997) Moreover, since there is no vaccine for HCV, HCV remains a serious public health problem.

A remarkable characteristic of HCV is its genetic heterogeneity, which is manifested throughout the genome (Bukh *et al.*, 1995). HCV circulates as a quasispecies of closely related genomes in an infected individual. Globally, six major HCV genotypes (genotypes 1-6) and multiple subtypes (a, b, c, etc.) have been identified (Bukh et al., 1993; Simmonds et al., 1993). The nucleotide and deduced amino acid sequences among isolates within a quasispecies generally differ by < 2%, whereas those between isolates of different genotypes vary by as much as 35%.

The most heterogeneous regions of the genome are found in the two envelope genes E1 and E2; in particular, the hypervariable region 1 (HVR1) at the N-terminus of E2 (Hijikata *et al.*, 1991; Weiner *et al.*, 1991).

The fact that the HVR1 is the region of the genome with the highest degree of genetic variability suggests that it is under strong immune pressure. Indeed, Ray et al. (1999) recently reported that patients who developed a chronic infection had a higher rate of non-synonymous mutations within the HVR1 as compared with the E1 protein while the reverse was observed in patients who were able to clear the infection. These authors therefore hypothesized that the HVR1 region of HCV might act as a decoy antigen by stimulating a strong immune response that is ineffective in clearing viremia.

However, the proposed role of the HVR1 as an immunologic decoy is not easily reconciled with prior studies in which a hyperimmune rabbit serum raised against the HVR1 was demonstrated to be capable of neutralizing HCV *in vitro* (Farci *et al.*, 1996, Shimizu *et al.*, 1996). Moreover, the presence of amino acids in the carboxy-terminal half of HVR1 which are

conserved across genotypes suggests that HVR1 might be required for HCV replication.

Summary of Invention

The present invention relates to nucleic acid molecules which lack all or part of the coding sequence of the HVR1 region of the envelope 2 (E2) gene of hepatitis C virus (HCV).

In the first embodiment, the nucleic acid molecule of the invention comprises the genome of an infectious hepatitis C virus in which the HVR1 of the envelope 2 gene of the infectious HCV has been deleted.

It is therefore an object of the invention to provide nucleic acid sequence which encodes infectious hepatitis C virus lacking HVR1. Such nucleic acid sequence is referred to throughout the application as "ΔHVR1-infectious nucleic acid sequence."

In a second embodiment, the nucleic acid molecule comprises chimeric genomes of chimeric hepatitis C viruses in which the structural region (core and envelope genes) or the envelope gene of a pestivirus genome, [for example, a bovine viral diarrhea virus (BVDV)] or a flavivirus genome (for example, a dengue virus or a yellow fever virus) are replaced by the corresponding structural region or E1 and E2 genes of an HCV in which the HVR1 region of HCV E2 has been removed. It is therefore an object of the invention to provide nucleic acid sequence which encodes chimeric HCV lacking the HVR1 of the HCV E2 gene. Such nucleic acid sequence is referred to throughout the application as "ΔHVR1-chimeric nucleic acid sequences."

The present invention also relates to the in vitro and in vivo production of ΔHVR1-infectious HCV or ΔHVR1-chimeric HCV

viruses from the ΔHVR1-infectious or ΔHVR1-chimeric nucleic acid sequences of the invention.

The present invention also relates to the use of the ΔHVR1 viruses of the invention to identify cell lines capable of supporting the replication of the viruses.

The invention also relates to the use of the ΔHVR1 infectious or chimeric nucleic acid sequences of the invention in the production of ΔHVR1-infectious HCV or ΔHVR1-chimeric HCV respectively, and the use of these virions for the development of inactivated or attenuated vaccines to prevent HCV in a mammal.

In a third embodiment, the DNA construct comprises an HCV E2 gene lacking the HVR1 region. Such a "ΔHVR1-E2 gene" may also be linked in tandem in the DNA construct with an HCV E1 gene. When contained alone or in tandem with the E1 gene, the ΔHVR1-E2 gene may be further modified at its carboxy-terminus to produce either a secreted or surface expressed ΔHVR1-E2 protein.

The invention further relates to pharmaceutical compositions and DNA-based vaccines which comprise the nucleic acid molecules of the invention.

The invention also relates to methods of preventing or treating HCV in a mammal comprising administering the nucleic acid molecules of the invention to a mammal in an amount effective to stimulate the production of a protective humoral and/or cellular immune response to HCV.

The invention also provides a kit for the treatment or prevention of HCV, the kit comprising a DNA molecule of the invention useful as an immunogen in generating a protective immune response to HCV.

The invention further relates to the use of the nucleic acid molecules of the invention as immunogens to

generate antibodies to the ΔHVR1-infectious HCV, the ΔHVR1-chimeric HCV or the ΔHVR1-E2 protein, preferably neutralizing antibodies. The invention therefore relates to the use of such antibodies in passive immunoprophylaxis and to pharmaceutical compositions which comprise these antibodies.

The invention also relates to transformation of host cells with nucleic acid molecules of the invention to produce host cells which express ΔHVR1-infectious HCV, the ΔHVR1-chimeric HCV or the ΔHVR1-E2 protein.

The invention further relates to the use of host cells expressing ΔHVR1-infectious HCV, the ΔHVR1-chimeric HCV or the ΔHVR1-E2 protein as immunogens to stimulate a protective immune response to HCV.

The present invention, of course, also relates to ΔHVR1 E2 protein produced from the ΔHVR1-E2 gene constructs of the invention or obtained from the ΔHVR1-infectious HCVs or ΔHVR1-chimeric HCVs of the invention. These ΔHVR1-E2 proteins may be used as vaccines for immunizing mammals, especially humans, against HCV.

Brief Description Of Figures

Figures 1A-1F show the nucleotide sequence (SEQ ID NO: 1) of the infectious hepatitis C virus clone of genotype 1a [H77C(ΔHVR1)] which lacks the hypervariable region one (HVR1) of the second envelope protein and Figures 1G-1H show the amino acid sequence (SEQ ID NO: 2) encoded by the clone. The complete sequence of H77C(ΔHVR1) is identical to pCV-H77C but lacks the fragment from nucleotide positions 1491 to 1571 which encodes HVR1 (Yanagi *et al.*, 1997; ATCC accession number PTA-157).

Figure 2 shows *in vitro* transcription-translation of constructs E1E2-715 (Lanes 1, 3, 4, 5, and 6) and ΔHVR1-7 (Lane 8) with rabbit reticulocyte lysates with and without the addition of canine microsomal membranes.

Figure 3 shows the results of qualitative reverse transcriptase-nested polymerase chain reaction (RT-PCR) for HCV-RNA, $\log_{10}$ HCV GE (genome equivalent) titer (in-house RT-PCR and Amplicor HCV Monitor, Roche Diagnostics), second generation ELISA for anti-HCV and serum levels of alanine aminotransferase (ALT) in chimpanzee 1590 following transfection with RNA transcripts of the HCV deletion mutant H77C(ΔHVR1).

Figure 4 shows infection of chimpanzee 96A008 with HCV lacking HVR1. Serum samples were collected weekly from the chimpanzee and monitored for HCV-RNA [in-house RT-nested PCR and HCV Monitor test version 2.0 (Roche)], HCV antibodies (second generation ELISA, Abbott Laboratories) and liver enzyme levels (ALT, Anilytics). PBMC were collected weekly and tested for HCV-specific proliferative capacity (peripheral CD4) with a panel of recombinant HCV proteins [C22 (core), C33-c (NS3), C100 (NS3-NS4), and NS5]. The peripheral CD8+ T cell response (peripheral CTL) was tested by stimulating PBMC with a large panel of HCV peptides corresponding to known CTL epitopes. Expanded T cells isolated from liver biopsy samples were tested for HCV-specific proliferative responses (Intrahepatic CD4) +, positive; - negative. Liver biopsies were examined also for necroinflammatory changes [0 (normal), 1+, 2+, 3+, 4+]. At week 0, the chimpanzee was inoculated intravenously with 90 ml of plasma from chimpanzee 1590 (week 4 after transfection). At a titer of 10 GE/ml this represented an inoculum of approximately 900 genome equivalents of HCV.

Detailed Description of The Invention

The present inventors surprisingly observed that HCV cDNA from which the HVR1 of the E2 gene had been deleted, encoded a hepatitis C virus which was able to replicate in vivo and stimulated a strong cellular immune response. Figure 1A-1F shows the nucleic acid sequence of this infectious HCV clone which was constructed using pCV-H77C (ATCC accession number PTA-157), an infectious HCV clone of genotype 1a.

The present invention therefore relates in one embodiment to a nucleic acid molecule which comprises the genome of an infectious hepatitis C virus in which the HVR1 of the envelope 2 gene of the infectious HCV has been deleted.

The E2 protein consists of amino acid 384 to 746 of the HCV polyprotein and the HVR1 of the E2 protein consists of amino acid 384 to 410.

It is therefore an object of the invention to provide nucleic acid sequence which encodes infectious hepatitis C virus lacking HVR1. Such nucleic acid sequence is referred to throughout the application as "ΔHVR1-infectious nucleic acid sequence."

It is understood that the deletion of HVR1 of the E2 gene in the HCV genome can be made in infectious HCV clones of any genotype. For example, infectious HCV clones of different genotypes which have been constructed include those of Kolykhalov et al., (1997) and Yanagi et al. (1997, 1998) [who reported the derivation from HCV strains H77 (genotype 1a) and HC-J4 (genotype 1b) of cDNA clones of HCV that are infectious for chimpanzees], and Yanagi et al. (Yanagi, 1999) [who reported the construction of an infectious cDNA clone from HCV strain HC-J6 (genotype 2a)]. Preferably, the infectious HCV clones are infectious HCV clones of genotype 1a (ATCC accession number PTA-

157), 1b (ATCC accession number 209596) or 2a (ATCC accession number PTA-153).

In a second embodiment, the nucleic acid molecule of the invention comprises the genome of a flavivirus or pestivirus in which one envelope gene of the flavivirus or pestivirus is replaced by the E2 gene of an infectious HCV from which the HVR1 region of HCV E2 has been removed.

It is therefore an object of the invention to provide nucleic acid sequence which encodes chimeric HCV lacking the HVR1 of the HCV E2 gene. Such nucleic acid sequence is referred to throughout the application as "ΔHVR1-chimeric nucleic acid sequence."

The *Flaviviridae* family of viruses which may be used to make the ΔHVR1-chimeric nucleic acid sequences include, but are not limited to, dengue virus, bovine viral diarrhea virus, yellow fever virus and Kunjin virus..

Preferably, the ΔHVR1-chimeric nucleic acid sequences of the invention are made using the structural region or E1 and E2 genes of an infectious HCV clone of any genotype. Preferably, the infectious HCV clones are infectious HCV clones of genotype 1a (ATCC accession number PTA-157), 1b (ATCC accession number 209596) or 2a (ATCC accession number PTA-153).

The deletion of HVR1 which may be made in the sequence of the invention is at least 5 amino acids in length, preferably 10 amino acids, and most preferably the entire 27 amino acids of HVR1. The nucleic acid molecules of the invention therefore comprise genomes of HCV which lack at least a fragment of the HVR1 sequence.

Of course, it is understood that in deleting all or part of the HVR1 sequence, one may extend the deletions further in the carboxy-terminal direction of the E2 protein as long as

the resultant nucleic acid molecule is capable of replicating in a chimpazee and producing the E2 protein lacking all or part of HVR1.

The ΔHVR1-infectious nucleic acid sequence of the invention may further include one or more mutations such as those described in Example 4 which result in amino acid changes. One mutation is located within the E2 gene and results in a change from leucine to histidine at amino acid position 615. Another mutation is located within the NS3 serine-protease domain and results in a change from arginine to histidine at amino acid position 1143. A third mutation is located within the NS5B RNA-polymerase domain and results in a change from glutamic acid to aspartic acid at amino acid position 2875.

The present invention further relates to the in vitro and in vivo production of hepatitis C viruses from the ΔHVR1-infectious nucleic acid sequences of the invention, and the production of chimeric viruses from the ΔHVR1-chimeric nucleic acid sequences of the invention.

In one embodiment, the ΔHVR1-infectious nucleic acid sequences or the ΔHVR1-chimeric nucleic acid sequences of the invention can be inserted into an expression vector that functions in eukaryotic cells. Eukaryotic expression vectors suitable for producing high efficiency gene transfer in vivo are well known to those of ordinary skill in the art and include, but are not limited to, plasmids, vaccinia viruses, retroviruses, adenoviruses and adeno-associated viruses.

The sequences contained in the recombinant expression vector can then be transcribed in vitro by methods known to those of ordinary skill in the art in order to produce RNA transcripts which encode the HCV or chimeric viruses of the invention. The HCV or chimeric viruses of the invention may

then be produced by transfecting cells by methods known to those of ordinary skill in the art with either the in vitro transcription mixture containing the RNA transcripts or with the recombinant expression vectors containing the nucleic acid sequences described herein.

Where transfection of cells with recombinant expression vectors containing the nucleic acid sequences of the invention is used, transfection may be done by methods known in the art such as electroporation, precipitation with DEAE-Dextran or calcium phosphate, or incorporation into liposomes.

Suitable cells or cell lines for culturing the HCV or chimeric viruses of the invention include, but are not limited to, EBTr and Huh7.

The present invention also relates to the use of the ΔHVR1-infectious nucleic acid sequences or the ΔHVR1-chimeric nucleic acid sequences of the invention to identify cell lines capable of supporting the replication of HCV and the chimeric viruses of the invention.

The invention further relates to the use of the ΔHVR1-infectious nucleic acid sequences or the ΔHVR1-chimeric nucleic acid sequences of the invention to develop inactivated or attenuated vaccines to prevent hepatitis C in a mammal. For example, virions from cell lines infected with the HCV or chimeric viruses of the invention, or transfected with a ΔHVR1-infectious nucleic acid sequence or a ΔHVR1-chimeric nucleic acid sequence of the invention, can be purified from the cells and inactivated by methods known to those of ordinary skill in the art. The inactivated virions can be used to immunize mice, and if neutralizing antibody to HCV is produced, the virions can then be used to immunize chimpanzees to determine whether the antibodies are protective. Alternatively, cells infected with

the viruses of the invention may be passaged in cell culture to produce attenuated viruses which can be tested as candidate live vaccines. In assaying the ability of the viruses of the invention to infect mammals one can assay sera or liver of the infected mammal by RT-PCR to determine viral titer. In addition, the virulence phenotype of the virus produced by transfection of mammals with the sequences of the invention can be monitored by methods known in the art such as measurement of liver enzyme levels (alanine aminotransferase (ALT) or isocitrate dehydrogenase (ICD)) or by histopathology of liver biopsies.

When used as a vaccine, the HCV or chimeric virions can be administered alone or in a suitable diluent, including, but not limited to, water, saline, or some type of buffered medium. The vaccine according to the present invention may be administered to an animal, especially a mammal, and most especially a human, by a variety of routes, including, but not limited to, intradermally, intramuscularly, subcutaneously, or in any combination thereof. Of course, it is understood that formulations or compositions comprising the HCV or chimeric virions of the invention may be used either therapeutically or prophylactically to treat or prevent the signs and symptoms of Hepatitis C.

In a third embodiment, the nucleic acid molecule of the invention comprises an HCV E2 gene lacking the HVR1 region. Such a "ΔHVR1-E2 gene" may also be linked in tandem in the DNA construct with an HCV E1 gene.

When contained alone or in tandem with the E1 gene, the ΔHVR1-E2 gene may be further modified at its carboxy-terminus to produce either a secreted or surface expressed ΔHVR1-E2 protein.

A ΔHVR1-E2 gene encoding E2 protein targeted to the cell surface is preferred. Such a construct may be constructed by fusing an endoplasmic reticulum signal sequence to the amino-terminus of the nucleic acid sequence which encodes the truncated E2 gene fused at its carboxy-terminus to a plasma membrane anchor sequence.

By endoplasmic reticulum (ER) signal sequence is meant a nucleic acid sequence which encodes a continuous stretch of amino acids, typically about 15 to about 25 residues in length, which are known in the art to be generally located at the amino terminus of proteins and are capable of targeting proteins to the endoplasmic reticulum. Such ER signal sequences are known to those of skill in the art (see, for example, van Heijne, G. J. Mol. Biol., (1985) 184:99-105) and those of skill in the art would understand that even though their amino acid sequences may vary, such ER signal sequences are functionally interchangeable. Examples of ER signal sequences which may be used in the chimeric genes of the invention include, but are not limited to, the 20-carboxy-terminal amino acids of the full-length HCV E1 protein (amino acids 364-383 of the HCV polyprotein), which serves as the natural signal sequence of the E2 protein or the murine Ig kappa-chain V-J2-C signal peptide sequence contained in the pDisplay vector.

Where the truncated HCV envelope protein is a truncated E2 protein, the approximately 30 carboxy-terminal amino acids of E2 have been identified to contain an ER retention sequence and its removal and replacement with a plasma membrane anchor sequence is believed to be critical for expression of the truncated E2 protein on the cell surface. Thus, the truncated E2 protein contains a truncation of at least the 20 carboxy-terminal amino acids of the full-length E2

protein, more preferably, a truncation of at least the 25 carboxy-terminal amino acids, and most preferably, a truncation of at least about the 30 carboxy-terminal amino acids. By "plasma membrane anchor sequence" as used in the chimeric gene of the invention is meant a nucleic acid sequence which encodes an amino acid sequence that allows for retention of at least part of the protein in the plasma membrane of a cell. At a minimum, a plasma membrane anchor sequence encodes a sequence of hydrophobic amino acids of sufficient length to span the lipid bilayer of the plasma membrane. Such hydrophobic sequences are known in the art as transmembrane domains and are typically found at the carboxy-terminus of many proteins found on the surface of cells or virions. These transmembrane domains are typically at least 20 to 30 amino acids in length and are followed by charged cytoplasmic domains of varying lengths.

It is therefore understood that the plasma membrane anchor sequence encoded by the coding sequence of the invention may contain in addition to a transmembrane domain of a virion or a protein found on the surface of a cell, a cytoplasmic domain.

Perferably, the encoded plasma membrane anchor sequence is at least twenty amino acids in length, more preferably, from about 20 to about 100 amino acids in length, and most preferably, from about 30 to about 70 amino acids in length. Examples of plasma membrane anchor sequences include, but are not limited to, hydrophobic transmembrane domains of receptors such as those for insulin and for a number of growth factors including platelet-derived growth factor (PDGF) and epidermal growth factor (EGF), as well as the transmembrane domains of viral proteins that are anchored in the lipid envelope of the intact virion such as the transmembrane domains of the vesicular stomatitis and rabies virus G proteins.

Preferred plasma membrane anchor sequences for inclusion in the chimeric genes of the invention are sequences which encode the 50 amino acid transmembrane domain of the PDGF receptor as contained in the pDisplay vector described in the Examples, the carboxy-terminal 64 and 37 amino acids respectively of the CD4 and decay accelerating factor (DAF) proteins (these sequences constitute the transmembrane and cytoplasmic domains of the CD4 and DAF proteins) and the 49 carboxy-terminal amino acids of the VSV G protein (also constituting the transmembrane and cytoplasmic domains of the VSV G protein).

Of course, one of ordinary skill in the art would readily understand that other transmembrane domains suitable for use as plasma membrane anchor sequences in the chimeric genes of the invention are known or could be readily identified by carrying out carboxy-terminal deletions of known plasma membrane or viral envelope proteins (see, for example, Men et al (J. Virol. (1991) 65; 1400-1407).

A ΔHVR1-E2 gene encoding a secreted E2 protein may be constructed by fusing an endoplasmic reticulum signal sequence to the nucleic acid sequence which encodes the truncated E2 gene. In this case, the truncated E2 gene contains a truncation of at least the 31 carboxy-terminal amino acids (amino acids 716-746), and more preferably, a truncation of at least the 85 carboxy-terminal amino acids (amino acids 662-746).

Like the ΔHVR1-infectious nucleic acid sequences and the ΔHVR1-chimeric nucleic acid sequences described above, it is understood that the ΔHVR1-E2 nucleic acid sequences of the invention may lack all or part of the HVR1 sequence. The deletion of HVR1 is at least 5 amino acids in length, preferably 10 amino acids, and most preferably the entire 27 amino acids of

HVR1. The ΔHVR1-E2 nucleic acid sequences of the invention may also contain further truncations at the carboxy-terminus of the E2 gene of HCV, and/or mutations at amino acid position 615 as described above.

The present invention therefore relates to insertion of the nucleic acid molecules comprising the ΔHVR1-E2 gene of the invention into a suitable expression vector that functions in eukaryotic cells, preferably in mammalian cells. By suitable it is meant that the vector is capable of carrying and expressing a chimeric gene of the invention. The expression vector therefore contains at least one promoter and any other sequences necessary or preferred for appropriate transcription and translation of the ΔHVR1-E2 gene. Preferred expression vectors include, but are not limited to, plasmid vectors.

The invention also relates to the use of expression vectors containing the ΔHVR1-E2 nucleic acid molecules of the invention as immunogens to produce protective antibodies to HCV. Direct transfer of the ΔHVR1-E2 nucleic acid sequences of the invention to a mammal, preferably a primate, more preferably a human, may be accomplished by injection by needle or by use of other DNA delivery devices such as the gene gun. Possible routes of administration of the expression vector include, but are not limited to, intravenous, intramuscular, intradermal, subcutaneous, intraperitoneal and intranasal.

Since the existence of different genotypes with a low degree of homology within the envelope proteins diminishes the hope of identifying conserved neutralization epitopes (Bukh J, et al, Sem Liver Dis (1995); 15:41-63), it is likely that a polyvalent vaccine will be needed to generate broadly reactive neutralizing antibodies. Thus, in a preferred embodiment, nucleic acid molecules comprising ΔHVR1-E2 nucleic acid

sequences of isolates from multiple genotypes of HCV may be administered together to provide protection against challenge with multiple genotypes of HCV.

Accordingly, those of ordinary skill in the art would readily understand that multiple copies of different ΔHVR1-E2 nucleic acid sequences may be inserted into a single vector such that a host cell transformed or transfected with the vector will produce multiple envelope proteins. For example, a polycistronic vector in which multiple different ΔHVR1-E2 genes may be expressed from a single vector is created by placing expression of each gene under control of an internal ribosomal entry site (IRES) (Molla, a. et al. Nature, 356:255-257 (1992); Gong, S.K. et al. J. of Virol., 263:1651-1660 (1989)). In a preferred embodiment, copies of different ΔHVR1-E2 nucleic acid sequences are inserted into multiple vectors and transformed or transfected into host cells so that multiple envelope proteins can be produced.

The expression vectors containing the ΔHVR1-E2 nucleic acid sequences of the invention may be supplied in the form of a kit, alone, or in the form of a pharmaceutical composition.

Suitable amounts of material to administer for prophylactic and therapeutic purposes will vary depending on the route selected and the immunogen (E2 gene with different deletions) administered. One skilled in the art will appreciate that the amounts to be administered for any particular treatment protocol can be readily determined without undue experimentation. The vaccines of the present invention may be administered once or periodically until a suitable titer of anti-HCV antibodies appear in the blood. A suitable amount of expression vector to be used for prophylactic purposes might be expected to fall in the range of from about 1 μg to about 5 mg,

more preferably from about 100 μg to about 5 mg, and most preferably from about 1 mg to about 2 mg. Such administration will, of course, occur prior to any sign of HCV infection. Further, one of skill in the art will readily understand that the amount of vector to be used will depend on the size and species of animal the vector is to be administered to.

A vaccine of the present invention may be employed in sterile liquid forms such as solutions or suspensions. Any inert carrier is preferably used, such as saline or phosphate-buffered saline, or any such carrier in which the expression vector of the present invention can be suitably suspended. The vaccines may be in the form of single dose preparations or in multi-dose flasks which can be utilized for mass-vaccination programs of both animals and humans. Of course, specific adjuvants such as CpG motifs (Krieg, A.K. et al.(1995) Nature 374:546 and Krieg et al. (1996)) J. Lab. Clin. Med., 128:128) may prove useful with DNA-based vaccines or other vaccines.

The DNA-based vaccines will normally exist as physically discrete units suitable as a unitary dosage for animals, especially mammals, and most especially humans, wherein each unit will contain a predetermined quantity of active material calculated to produce the desired immunogenic effect in association with the required diluent. The dose of said vaccine or inoculum according to the present invention is administered at least once. In order to increase the antibody level, a second or booster dose may be administered at some time after the initial dose. The need for, and timing of, such booster dose will, of course, be determined within the sound judgment of the administrator of such vaccine or inoculum and according to sound principles well known in the art. For example, such booster dose could reasonably be expected to be advantageous at

some time between about 2 weeks to about 6 months following the initial vaccination. Subsequent doses may be administered as indicated.

The ΔHVR1-E2 nucleic acid sequences of the present invention can also be administered for purposes of therapy, where a mammal, especially a primate, and most especially a human, is already infected, as shown by well-known diagnostic measures. When expression vectors containing the chimeric genes of the present invention are used for such therapeutic purposes, much of the same criteria will apply as when they are used as a vaccine, except that inoculation will occur post-infection. Thus, when the expression vectors of the present invention are used as therapeutic agents in the treatment of infection, the therapeutic agent comprises a pharmaceutical composition containing a sufficient amount of the expression vector so as to elicit a therapeutically effective response in the organism to be treated. Of course, the amount of pharmaceutical composition to be administered will, as for vaccines, vary depending on the immunogen contained therein and on the route of administration.

The therapeutic agent according to the present invention can thus be administered by subcutaneous, intramuscular, intradermal or intranasal routes. One skilled in the art will certainly appreciate that the amounts to be administered for any particular treatment protocol can be readily determined without undue experimentation. Of course, the actual amounts will vary depending on the route of administration as well as the sex, age, and clinical status of the subject which, in the case of human patients, is to be determined with the sound judgment of the clinician.

The therapeutic agent of the present invention can be employed in sterile liquid forms such as solutions or

suspensions. An inert carrier is preferably used, such as saline, phosphate-buffered saline, or any such carrier in which the expression vectors of the present invention can be suitably suspended. The therapeutic agents may be in the form of single dose preparations or in multi-dose flasks, which can be utilized for mass-treatment programs of both animals and humans. Of course, when the expression vectors of the present invention are used as therapeutic agents, they may be administered as a single dose or as a series of doses, depending on the situation as determined by the person conducting the treatment.

The invention also relates to transformation of host cells with nucleic acid molecules of the invention to produce host cells which express HVR1-infectious HCV, the ΔHVR1-chimeric HCV or the ΔHVR1-E2 protein.

The invention further relates to the use of host cells expressing ΔHVR1-infectious HCV, the ΔHVR1-chimeric HCV or the ΔHVR1-E2 protein as immunogens to stimulate a protective immune response to HCV.

The present invention, of course, also relates to ΔHVR1 E2 protein produced from the ΔHVR1-E2 gene constructs of the invention or obtained from the ΔHVR1-infectious HCVs or ΔHVR1-chimeric HCVs of the invention.

The nucleic acids, polypeptides and viruses of the present invention can also be utilized in the production of antibodies against HCV. The term "antibody" is herein used to refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules. Examples of antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and portions of an immunoglobulin molecule, including those portions known in the

art as Fab, $F(ab')_2$ and F(v) as well as chimeric antibody molecules.

Thus, the polypeptides, viruses and nucleic acid sequences of the present invention can be used in the generation of antibodies that immunoreact (i.e., specific binding between an antigenic determinant-containing molecule and a molecule containing an antibody combining site such as a whole antibody molecule or an active portion thereof) with antigenic determinants on the surface of hepatitis C virus particles.

The present invention therefore also relates to antibodies produced following immunization with the nucleic acid sequences, viruses or polypeptides of the present invention. These antibodies are typically produced by immunizing a mammal with an immunogen or vaccine to induce antibody molecules having immunospecificity for polypeptides or viruses produced in response to infection with the nucleic acid sequences of the present invention. When used in generating such antibodies, the nucleic acid sequences, viruses, or polypeptides of the present invention may be linked to some type of carrier molecule. The resulting antibody molecules are then collected from said mammal. Antibodies produced according to the present invention have the unique advantage of being generated in response to authentic, functional polypeptides produced according to the actual cloned HCV genome.

The antibody molecules of the present invention may be polyclonal or monoclonal. Monoclonal antibodies are readily produced by methods well known in the art. Portions of immunoglobin molecules, such as Fabs, as well as chimeric antibodies, may also be produced by methods well known to those of ordinary skill in the art of generating such antibodies.

The antibodies according to the present invention may also be contained in blood, plasma, serum, hybridoma supernatants, and the like. Alternatively, the antibody of the present invention is isolated to the extent desired by well known techniques such as, for example, using DEAE Sephadex. The antibodies produced according to the present invention may be further purified so as to obtain specific classes or subclasses of antibody such as IgM, IgG, IgA, and the like. Antibodies of the IgG class are preferred for purposes of passive protection.

The antibodies of the present invention are useful in the prevention and treatment of diseases caused by hepatitis C virus in animals, especially mammals, and most especially humans.

In providing the antibodies of the present invention to a recipient mammal, preferably a human, the dosage of administered antibodies will vary depending on such factors as the mammal's age, weight, height, sex, general medical condition, previous medical history, and the like.

In general, it will be advantageous to provide the recipient mammal with a dosage of antibodies in the range of from about 1 mg/kg body weight to about 10 mg/kg body weight of the mammal, although a lower or higher dose may be administered if found desirable. Such antibodies will normally be administered by intravenous or intramuscular route as an inoculum. The antibodies of the present invention are intended to be provided to the recipient subject in an amount sufficient to prevent, lessen or attenuate the severity, extent or duration of any existing infection.

The invention also provides that the nucleic acid molecules, viruses, polypeptides and antibodies of the invention may be supplied in the form of a kit, alone or in the form of a pharmaceutical composition.

All scientific publication and/or patents cited herein are specifically incorporated by reference. The following examples illustrate various aspects of the invention but are in no way intended to limit the scope thereof.

EXAMPLES

Materials and Methods

Construction of An Expression Vector ΔHVR1-7 Containing E1 and a Carboxy-Terminal Truncated E2 Which Lacks the HVR1 Region and the Carboxy-Terminal 31 Amino Acids of E2

Deletion of the nucleotide sequence encoding the HVR1 region was performed by fusion PCR using pCV-H77C, the infectious cDNA clone of HCV genotype 1a (Yanagi *et al.*, 1997;ATCC accession number PTA-157).

Table 1

Primers used for the construction of ΔHVR1*

PRIMER	SEQUENCE
SEQ: ID NO:3	
E2-PstI:	5'-ACG CGT CTG CAG CTT AAT GGC CCA GGA CGC GAT GCT TG-3'
SEQ: ID NO:4	
E1-BglII:	5'- ACG CGT AGA TCT TAC CAA GTG CGC AAT TCC TCG GGG-3'
SEQ: ID NO:5	
E1-383R (+10):	5'-TCA GTT GGA TAG CGT CGA CGC CGG CAA ATA GCA G-3'
SEQ: ID NO:6	
E2-411S (+10):	5'-CGT CGA CGC GAT CCA ACT GAT CAA CAC CAA CGG C-3'
SEQ: ID NO:7	
HVR1-mutS:	5'-CTT GTA CCA TCA ATT ACA CCA TAT TC-3'
SEQ: ID NO:8	
HVR1-mutR:	5'-GAT AGT GCC AAT GCC TAT ACG GG-3'
SEQ: ID NO:9	
FUSION 1:	5'-CGT ATA GGC ATT GGC ACT ATC CTT GTA CCA TCA ATT ACA CC-3'
SEQ: ID NO:10	
FUSION2:	5'-GGT GTA ATT GAT GGT ACA AGG ATA GTG CCA ATG CCT ATA CG-3'
SEQ: ID NO:11	
E2-364 NotI:	5'-TTT TTT TTG CGG CCG CAT GGT GGG GAA CTG GGC GAA GGT CC-3'
SEQ: ID NO:12	
E2-661-HindIII:	5'-ACG CGT AAG CTT CTA TTA CTC GGA CCT GTC CCT GTC TTC CAG-3'

*Restriction sites within primers are underlined

In the first step, the sequences encoding the E1 protein (aa 192-383) and a truncated E2 protein (aa 411-715) lacking the HVR1 region (aa 384-410) were PCR amplified from pCV-H77C. Briefly, 50 ng of cDNA were added to a master mix containing 4 µl of 5X Advantage KlenTaq Buffer, 1.25 µl of dNTP (10 mM), 1 µl of 10 µM sense primer [E1-BglII (SEQ ID NO: 4) for E1 and E2-411S(+10) (SEQ ID NO: 6) for E2], 1 µl of 10µM antisense primer [E1383R (+10) for E1 and E2-PstI for E2] and 1 µl of the Advantage KlenTaq polymerase mixture. Cycling conditions were 99°C for 1 min followed by 25 cycles of 99°C for 35 sec, 67°C for 30 sec, 68°C for 3 min 30 sec. After purification of the PCR products, a fusion PCR was performed. Briefly, 2.5 µl of each PCR product were added to a master mix containing 10 µl of 10X Pfu buffer, 2 µl of 10 mM dNTP, 5 µl of 10 µM sense primer (E1-BglII)(SEQ ID NO: 4), 5 µl of 10 µM antisense primer (E2-PstI)(SEQ ID NO: 3), 1 µl of Pfu and 72 µl of H_2O. Cycling conditions were 95 °C for 1 min, 67 °C for 1 min, 72 °C for 3 min 30 sec (30 cycles). The fusion PCR product was digested with *Bgl*II and *Pst*I and cloned into the expression vector pDisplay (Invitrogen). A clone with an insert of the correct size (ΔHVR1-7) was selected and sequenced. Sequence analysis of both strands of DNA confirmed that this clone contained the expected sequence of E1 and a truncated E2 lacking the HVR1, i.e., ΔHVR1-7 encodes amino acids 192-383 of the E1 protein fused to amino acids 411-715 of the E2 protein.

Construction of An Infectious clone of HCV Lacking HVR1

Construction of an infectious cDNA clone that was full-length except for the HVR1 was performed by digestion of the expression vector ΔHVR1-7 with *Mun*I (which cuts at

nucleotide positions 1254 and 1983 of pCV-H77C) and cloning of the resulting fragment into the digested HCV cDNA clone H77C. Briefly, a 3 µg sample of the infectious HCV cDNA clone H77C was digested with *Mun*I for 3 hours and the enzyme was then inactivated at 65°C for 20 min. The digested cDNA clone, lacking a portion of the E1 and E2 regions (nucleotides 1255 to 1983 of pCV-H77C), was dephosphorylated with calf intestinal alkaline phosphatase. A sample of ΔHVR1-7 was then digested with *Mun*I for 3 hours. After inactivation of the *Mun*I enzyme, the digested fragment was ligated into the digested pCV-H77C using standard procedures (Forns *et al.*, 1999). A clone containing the correct insert was selected, retransformed and large-scale plasmid DNA was prepared as previously described (Yanagi *et al.*, 1997). The complete sequence of the HVR1 deletion mutant [H77C(ΔHVR1)] was the expected one, that is, identical to pCV-H77C but lacking the fragment from nucleotide positions 1491 to 1571 which encodes HVR1.

In vitro transcription-translation analysis

Non-linearized plasmid ΔHVR1-7 and the control plasmid E1E2-715 were used for *in vitro* transcription-translation. The ΔHVR1-7 plasmid contained the nucleotide sequence encoding E1 (aa 192-383) and E2 (aa 411-715) inserted between a leader sequence which targeted the HCV proteins to the secretory pathway and the transmembrane domain of the platelet-derived growth factor receptor (PDGFR) which anchored the HCV E2 proteins to the plasma membrane in the expression vector pDisplay (Invitrogen). The control plasmid contained the nucleotide sequence encoding E1 (aa 192-383) a carboxy-truncated form of E2 (aa 384-715) of pCV-H77C inserted between a leader sequence which targeted the HCV proteins to the secretory pathway and the transmembrane domain of PDGFR which anchored the

HCV E2 proteins to the plasma membrane in the expression vector pDisplay. Reactions were performed in 25 µl of the TNT Coupled Reticulocyte Lysate System (Promega) containing [S^{35}] methionine, with or without the addition of canine microsomal membranes at 30 °C for 90 min. Total translation products were separated in 12% SDS/PAGE and identified by autoradiography.

Analysis of cell surface-expression of the E2 protein

Huh7 cells grown in 4-well tissue culture chambers were transfected with plasmids E1E2-715 and ΔHVR1-7 described above. Immuno-fluorescence analysis was performed 48 hours after transfection. Live cells were incubated for 30 minutes with a 1:100 dilution of rabbit hyperimmune serum (*Lmf86*) raised against a peptide encoding the carboxy-terminal 21 amino acids of HVR1 (aa 390-410 of pCV-H77C) or a rabbit hyperimmune serum (*FOR-1*) raised against a peptide within E2, but outside the HVR1 (aa 517-535 of pCV-H77C). After washing, cells were incubated with FITC-conjugated goat anti-rabbit IgG (1:100) for 30 min at 37°C. After washing, slides were mounted and examined for immunofluorescence. For intracellular immunofluorescence staining , cells were fixed and permeabilized with cold acetone, as previously described (Forns *et al.*, 1999).

Transfection of a chimpanzee with H77C(ΔHVR1) RNA

RNA was transcribed *in vitro* with T7 RNA polymerase from 10 µg of template plasmid H77C(ΔHVR1) linearized with *XbaI* as described previously (Yanagi et al., 1997). The quality and amount of RNA were analyzed by gel electrophoresis and ethidium bromide staining. Two transcription mixtures were each diluted with 400 µl of ice-cold phosphate-buffered saline without calcium or magnesium and then immediately frozen on dry ice and stored at -80°C. Within 24 hours, both transcription mixtures

were percutaneously injected into the liver of a chimpanzee (number 1590) under ultrasonographic control. The housing, maintenance, and care of the animals were in compliance with all relevant guidelines and requirements.

Serum samples were collected weekly from the chimpanzee and monitored for serum levels of alanine aminotransferase (ALT), anti-HCV antibodies [second generation ELISA] and HCV-RNA [HCV Monitor test (Roche) and in-house RT-nested PCR (Bukh *et al.*, 1998)]. Sequence analysis of the recovered virus was performed at different time points during follow-up. In short, genomic regions were amplified in RT-nested PCR (Bukh *et al.*, 1998) with primers specific for the H77 strain of HCV.

Construction of vectors that Express Mutant E2 Protein on the Cell Surface or in a Form That Is Secreted

A. Cell Surface ΔHVR1-E2 Vectors

Clones E1E2-715 (amino acids 192-715) and ΔHVR1-7 (amino acids 192-383 of E1 in frame with amino acids 411-715 of E2) were produced as described above.

Clone ΔHVR1-mut5 (amino acids 192-383 of E1 in frame with amino acids 411-715 of E2 except that amino acid 615 is changed from a leucine to a histidine was produced as follows.

Primers E1-BglII (SEQ ID NO: 4) and HVR1-mutR (SEQ ID NO: 8)(containing a T->A mutation at nucleotide position 2185: see Example 4) were used to amplify a fragment encoding the E1 protein and a portion of the E2 protein using Advantage KlenTaq Polymerase mix, as described above. Primers HVR-1 mutS (SEQ ID NO: 7) and E2-PstI (SEQ ID NO: 3) were used to amplify the rest of the truncated E2 protein by using AmpliTaq Gold DNA polymerase. In a second step, both products were fused by an

amplification reaction that contained a mixture of both external primers (E1-BglII (SEQ ID NO: 4) and E2-PstI (SEQ ID NO: 3) and the two fusion primers SEQ ID NO: 9 and SEQ ID NO: 10)(Table 1), by using Advantage KlenTaq Polymerase mix. The fusion product was digested with *Bgl*II and *Pst*I and cloned into the digested pDisplay vector to produce ΔHVR1-mut5.

Clones E1E2-715, ΔHVR1-7 and ΔHVR1-mut5 were used to transfect Huh7 cells and to determine their pattern of reactivity against a panel of 5 rabbit hyperimmune sera and 12 human monoclonal antibodies (Cardoso *et al.*, 1998, Inchauspe et al. 1998) by immunofluorescence in live and fixed/permeabilized cells.

B. Secreted ΔHVR1-E2 Vectors

In addition, three different expression vectors encoding a secreted form of the E2 protein were constructed: E2-661, encoding the ER signal sequence of E1 (aa 364-383 of pCV-H77C) and a carboxy-terminal truncated E2 protein (aa 384-661 of pCV-H77C); E2-661 ΔHVR1, encoding the same protein with a deletion of the HVR1 region (aa 384-410 of pCV-H77C); and E2-661 ΔHVR1-mut, encoding the same protein with a deletion of the HVR1 and the replacement of leucine for histidine at amino acid position 615. In this case, amplification was performed from templates E1E2-715, ΔHVR1-7 and ΔHVR1-mut5, respectively, with primers shown in Table 1. PCR products were cloned into the expression vector pcDNA3.1(-) (Invitrogen). Clones E2-661, E2-661 ΔHVR1, and E2-661 ΔHVR1-mut were used to express soluble E2 protein *in vitro*. Briefly, one microgram of each of the plasmids was used for *in vitro* transcription-translation. Reactions were performed in 25 µl of the TNT Coupled Reticulocyte Lysate System (Promega) containing [S^{35}] methionine,

with the addition of canine microsomal membranes at 30°C for 90 min.

Example 1

In vitro transcription-translation analysis

To determine whether cleavage between E1 and E2 was affected by the deletion of the HVR1 region, expression vectors E1E2-715 (encoding the complete E1 protein and a cell-surface E2 protein lacking the carboxy-terminal 31 amino acids of E2) and ΔHVR1-7 (encoding the complete E1 protein and a cell-surface E2 lacking the HVR1 and carboxy-terminus 31 amino acids) were assayed for *in vitro* protein synthesis. *In vitro* transcription-translation was performed with the TNT Coupled Reticulolysate System using [S^{35}]methionine. In the absence of microsomal membranes, translation products of the expected size were obtained for E1E2-715 and ΔHVR1-7 (Figure 2). When microsomal membranes were added, cleavage between E1 and E2 was observed in both cases (Figure 2) and, as expected, the E2 protein derived from ΔHVR1-7 was slightly smaller in size. These results indicate that cleavage between E1 and E2 is not affected by the deletion of the N-terminus (HVR1) of the E2 protein.

Example 2

Analysis of cell-surface expression of the E2 protein

To analyze whether a truncated E2 protein lacking the HVR1 region could be targeted to the cell surface, Huh7 cells were transfected with E1E2-715 and ΔHVR1-7. After transfection, fixed/permeabilized cells and live cells were analyzed for expression of E2 by immunofluorescence. Up to 20% of cells transfected with E1E2-715 were positive when stained with a rabbit hyperimmune serum *lmf 86* or *FOR-1*. Both fixed

/permeabilized and live cells were stained. A similar proportion of cells transfected with ΔHVR1-7 was positive when stained with rabbit hyperimmune serum *FOR-1* (raised against aa 517-535). Again, both fixed/permeabilized and live cells were stained. In contrast, there was no staining with rabbit hyperimmune serum *lmf86,* which was raised against aa 390-410 (within HVR1). These results demonstrated that transport of the E2 protein to the cell surface was not affected by deletion of the HVR1 region, suggesting that deletion of the N-terminal domain of E2 did not result in misfolding of this protein.

Example 3

Transfection of a chimpanzee with RNA of H77C(ΔHVR1)

To determine the infectivity of H77C(ΔHVR1), RNA transcripts were injected into the liver of chimpanzee 1590. HCV-RNA was detected by RT-nested PCR at week one post-inoculation (p.i.) and the animal remained HCV-RNA positive throughout the entire follow-up period of 22 weeks (Figure 3). The quantitative HCV Monitor test was negative until week 10 p.i. The genome titer during the first 7 weeks was □ 10 genome equivalent (GE)/ml. However, beginning at week 8, the titer progressively increased over time (Table 2 and Figure 3).

Table 2

Chimpanzee 1590 Infected with H77C(ΔHVR1)

Week	HCV RNA*	Titer† in house	Monitor‡	E2 (nt 2185)	NS3 (nt 3769)	NS5B (nt 8966)
0	-					
1	+/-					
2	+/-					
3	+	1				
4	+	1		T	G	
5	+	1		T	G	A
6	+	1		T	G	A
7	+	1	-	T	G	A
8	+	2	-	T/A	G	A/t
9	+	2	-	T/A	G/a	T/a
10	+	2	119	A/t	G/a	T/a
11	+	3	756	A	G/a	T
12	+	3	682	A	G/A	T
13	+	4	1981	A	A	T
14	+	4	3128			
15	+		3038			
16	+		3285			
17	+		1928			
18	+		5292			

* In-house RT-PCR qualitative test
† log10 HCV GE titer by in-house RT-PCR
‡ Amplicor HCV monitor (number of genome copies/ml) (Roche Diagnostics)
Nucleotide sequence corresponding to Yanagi 1997
Note: Lower case letters represent minor species

Serum liver enzyme values remained normal during the follow-up. The second generation anti-HCV test remained negative at least through week 22 p.i.

The follow-up period for chimpanzee 1590 has been extended to 72 weeks. Chimpanzee 1590 remained HCV-RNA positive throughout the 72 week follow-up period. Thus, the chimpanzee developed a chronic infection. In addition, the genome titer reached 10^3 - 10^4 GE/ml at weeks 11-14 (see Table 2) and remained at 10^3 - 10^4 GE/ml throughout the remainder of the 72 week period.

At weeks 51 and 52 p.i., the relative genome titer of HCV lacking HVR1 was also assessed in serum, PMBC, and liver obtained from the transfected chimpanzee. HCV RNA was not

detected in PBMCs, but the HCV titers determined from liver tissue at weeks 51 and 52 p.i. were higher than the titers determined from the serum.

The 72 week follow-up period also revealed that the transfected chimpanzee became positive for antibodies to second generation ELISA HCV at week 37 and third generation RIBA (Chiron) confirmed the presence of antibodies to C22(core), C33-c (NS3), and c100 (NS3-NS-4).

Finally, CD4+ and CD8+ T cell responses were also measured. An HCV-specific proliferative CD4+ T cell response to C22 (core), C33-c (NS3), c100(NS3-NS4) and/or NS5 antigens of HCV was detected in the peripheral blood mononuclear cells (PBMC) beginning at week one post-infection and throughout the follow-up period. During most weeks, however, only a mono-specific response to the core was detected.

Peripheral CD8+ T cell (CTL) responses were tested at weeks 5, 9, 14, and 18 by an *in vitro* peptide stimulation assay. CTL were detected with two NS4 peptides at weeks 9 and 14 and with one of these at week 18. A transient and monospecific (NS5) proliferative T cell response was detected in the liver at weeks 14 and 15. Although serum liver enzyme values remained normal during the entire follow-up, necroinflammatory changes indicative of hepatitis were detected in liver biopsies during weeks 34-40.

Example 4

Analysis of the HCV-RNA recovered from chimpanzee 1590

Sequence analysis of the HCV-RNA recovered from chimpanzee 1590 was performed at week 13. The entire ORF was amplified by RT-nested PCR of overlapping gene regions. Direct sequence analysis of both strands of DNA from the PCR products

confirmed that the recovered virus lacked the HVR1 region. Interestingly, three nucleotide mutations were identified; all three mutations were non-synonymous i.e. resulted in an amino acid change (Table 2). One mutation was located within the E2 region (corresponding to aa position 615 of H77C) and generated a change from leucine to histidine. Another mutation was located within the serine-protease domain of NS3 (corresponding to aa position 1143 of H77C; arginine to histidine). Both mutations are at positions that are universally conserved among all HCV genotype reference strains (Bukh *et al.*, 1998). Finally, a third mutation was observed in the NS5B RNA-polymerase domain (corresponding to aa position 2875 of H77C; glutamic acid to aspartic acid). Although this is not a universally conserved position, a substitution of aspartic acid has not been described previously. Thus, all three mutations were unique.

Two additional mutations were identified upon rereading of the week 13 sequence: mutation at 514 (valine to methionine) and a synonymous mutation at amino acid 2631. Sequence analysis of the HCV-RNA recovered from chimpanzee 1590 was also performed at week 24 and five nucleotide mutations were identified; four of which were non-synonymous. (Table 3).

Table 3

Evolution of HCV lacking HVR1 in chimpanzee 1590*. We performed sequence analysis of the entire open reading frame of the virus recovered from chimpanzee 1590 at weeks 13 and 24 post-transfection. We confirmed that the HVR1 region was not present and identified differences from the wild-type cDNA clone as indicated.

			E2				NS3				NS5B			
Week	Titer In-house†	Titer Monitor‡	nt.1881	aa.514	nt.2185	aa.615	nt.3769	aa.1143	nt.4708	aa.1456	nt.8234	aa.2631	nt.8966	aa.2875
			G	V	T	L	G	R	C	T	G	K	A	E
1	<1	Neg												
2	<1	Neg												
3	1	Neg												
4	1	Neg			T	L	G	R						
5	1	Neg	G	V	T	L	G	R					A	E
6	1	Neg	G	V	T	L	G	R					A	E
7	1	Neg	G	V	T	L	G	R					A	E
8	2	Neg	G	V	T/A	L/H	G	R					A/t	E/d
9	2	Neg	G	V	T/A	L/H	G/a	R/h					T/a	D/e
10	2	2.08	G	V	A/t	H/l	G/a	R/h					T/a	D/e
11	3	2.88	G	V	A	H	G/a	R/h					T	D
12	3	2.83	G/a	V/m	A	H	G/A	R/H					T	D
13	4	3.30	G/A	V/M	A	H	G/A	R/H	C	T	G/a	K	T	D
14	4	3.50	G/A	V/M			G/A	R/H	C	T	G	K		
15	3	3.48												
16	4	3.52	A/g	M/v			G/a	R/h	C	T	G/a	K		
17	4	3.29												
18	3	3.72	A	M			G	R	C/t	T/m	G/A	K		
19	3	3.20												
20	3	3.37	A	M			G	R	C/T	T/M	A/g	K		
21	3	4.04												
22	4	3.91	A	M			G	R	C/T	T/M	A/g	K		
23	4	3.48												
24	4	3.74	A	M	A	H	G	R	T/c	M/t	A	K	T	D

*Nucleotide (nt) and amino acid (aa) positions of H77C . Sequence of H77C(ΔHVR1) cDNA is shown on top. Dominant sequences recovered from the chimpanzee are shown in capital letters. Minor sequences recovered from the chimpanzee are shown in lower case letters. †$\log_{10}$ titer determined by RT-nested PCR on 10-fold serially diluted extracted RNA. ‡$\log_{10}$ titer determined by second generation Monitor test (Roche).

Four of the mutations were identical to the mutations identified at week 13 and one of the week 13 mutations (the NS3 mutation at amino acid 1143) reverted to wild-type. The single new mutation identified at week 24 was located within the serine protease domain of NS3 (corresponding to aa position 1456 of H77C; threonine to methionine).

To analyze a possible relationship between the increase in viral titer and the appearance of the four non-synonymous mutations identified at week 13, sequence analysis was performed from regions encompassing the described mutations. During weeks 4-7, direct sequence analysis did not reveal evidence of the mutated nucleotides at the four positions. Between weeks 8 and 14, coinciding with the first increase in viral titer, direct sequence analysis revealed the presence of amino acid substitutions at four positions (two in E2, one in NS3, and one in NS5B)(See Table 3). Therefore, there was a temporal association between the appearance of these mutations and an increase in viral titer (See Tables 2 and 3).

Example 5

Transmission of HCV lacking HVR1 to a second chimpanzee

To investigate whether the HCV mutant lacking HVR1 was transmissible, a naïve chimpanzee (#96A008) was inoculated intravenously with 90 ml of plasma taken at week 4 from chimpanzee 1590, before appearance of consenus mutations in that chimpanzee (Table 3). The virological and immunological correlates of HCV infection in chimpanzee 96A009 are shown in Fig. 4. Serum HCV-RNA was first detected at week 2 p.i., with a genome titer of 10^1 GE/ml and the titer increased to peak levels of 10^4-10^5 GE/ml during weeks 3-9. The infection was resolved at week 18 p.i. Antibodies to HCV were not detected. A peripheral

mono-specific proliferative CD4+ T cell response against core (C22) was detected during weeks 4-13 p.i.; a multi-specific response against NS3 (C33-c), NS3-NS4 (c100) and NS5 was detected at week 18. The chimpanzee mounted an early and multi-specific peripheral CD8+ T cell response to a total of seven different epitopes (representing core, NS3, NS4 and NS5) at the weeks tested (weeks 4, 7, 11, 15 and 19). Finally, a multi-specific (C33-c, c100 and NS5) sustained proliferative CD4+ T cell response was detected in the liver from week 8 p.i. Liver enzyme values were marginally elevated during weeks 10-18 p.i. and necroinflammatory changes were detected in liver biopsies during weeks 11-19 p.i.

Sequence analysis of the entire open reading frame of HCV recovered from the chimpanzee 96A008 at weeks 4 and 9 p.i. showed that the transmitted virus lacked HVR1. Compared with the cDNA clone of that mutant, four nucleotide substitutions were identified in the virus and all four mutations were non-synonymous. (Table 4).

Table 4

Evolution of HCV lacking HVR1 in chimpanzee 96A008*. We performed sequence analysis of the entire open reading frame of the virus recovered from chimpanzee 96A008 at weeks 4 and 9 post-inoculation. We confirmed that the HVR1 region was not present and identified differences from the wild-type cDNA clone as indicated.

			E1		E2		NS3			
Week	Titer In-house†	Titer Monitor‡	nt.1097	aa.252	nt.1881	aa.514	nt.3769	aa.1143	nt.4269	aa.1310
			A	K	G	V	G	R	T	Y
1	Neg	Neg								
2	1	Neg	A/T	K/N	A	M	G	R	C	H
3	3	3.33	T	N	A	M	G/a	R/h	C	H
4	4	4.00	T	N	A	M	G/a	R/h	C	H
5	4	3.98					G/A	R/H		
6	5	4.25					G/A	R/H		
7	5	4.81					G/A	R/H		
8	5	4.51					G/A	R/H		
9	5	4.62	T	N	A	M	G/A	R/H	C	H
10	4	3.30					G/a	R/h		
11	3	2.63								
12	3	2.92					G	R		
13	2	Neg								
14	2	3.14					G	R		
15	1	Neg								
16	4	3.06					G	R		
17	1	Neg								
18	Neg	Neg								
19	Neg	Neg								
20	Neg	Neg								

*Nucleotide (nt) and amino acid (aa) positions of H77C. Sequence of H77C(ΔHVR1) cDNA is shown on top. Dominant sequences recovered from the chimpanzee are shown in capital letters. Minor sequences recovered from the chimpanzee are shown in lower case letters. †$\log_{10}$ titer determined by RT-nested PCR on 10-fold serially diluted extracted RNA. ‡$\log_{10}$ titer determined by second generation Monitor test (Roche).

Sequence analysis of regions encompassing the described mutations showed that a mutation in E2 and NS3 each had replaced the wild-type nucleotide by the time the virus was first detected (week 2). An identical change in E2 had occurred in the transfected chimpanzee (Table 3). A change in E1 appeared as the quasispeicies at week 2 but had replaced the wild-type nucleotide completely by week 3. The other mutation, in NS3, appeared at week 3 and persisted as a quasispecies during weeks 3-10 and then reverted to wild-type. This latter mutation had been seen also in the transfected chimpanzee. (Table 3).

Example 6

Analysis of potential conformational changes within E2 after L615H substitution

Sequence analysis of HCV recovered from chimpanzee 1590 demonstrated that a non-synonymous mutation occurred within the E2 protein during follow-up (nt 2186 of pCV-H77C; amino acid position 615). Since a change from a non-polar amino acid (leucine) to a charged one (histidine) within E2 might change its conformation, Huh7 cells were transfected with E1E2-715 (intact E1-E2), ΔHVR1-7 (the deletion of HVR1) or ΔHVR1-mut (deletion of HVR1 along with the mutation at position 615), and the transfected cells were stained for E2 with a panel of 4 rabbit hyperimmune sera and 12 human monoclonal antibodies (Cardoso *et al.* 1998; Inschauspe *et al.*, 1998). Immunofluoresecence analysis showed that there was no significant change in the patterns of reactivity among the three different forms of E2 targeted to the cell surface (Table 5).

Table 5

Effect of deletion of the HVR1 region and mutation of E2 615 (leu-his) on the reactivity of cell surface-expressed E2

Antibody	E1E2-715	ΔHVR1-7	ΔHVR1-mut5
RHS *1mf86* (aa390-410)	++++	-	-
RHS 521 (aa 645-662)	++	+++	++
RHS FOR-1(aa 517-535)	+++	+++	+++
RHS 554 (aa 460-483)	++++	+++	+++
2H1	+/-	-	-
2G1	++	++	+
1H3	++	+++	+
4F7	++	++	+
8E8	-	-	-
2H8	+++	+++	++
4E5	++	+++	++
4F1	++	+++	+++
7B7	+++	++	++
108F6	++	+	++
503E7p	+++	++	+
29E4p	++	++	+

-RHS: rabbit hyperimmune sera.
-Amino acid position within the H77C sequence of the peptides used to generate RHS are indicated in parenthesis.
-The remaining 12 antibodies are human monoclonal antibodies (Cardoso *et al.* 1998; Inschauspe *et al.*, 1998).
-Scoring system from - to ++++. - means no staining and ++++ means strong staining.

References

Alter, M. J. (1997). Hepatology **26**, 62S-65S.

Bukh, J., Miller, R.H., and Purcell, R.H. (1995) *Semin. Liver Dis.* 15, 41-63

Bukh, J., *et al.* (1998) *Hepatology* 28, 319A

Bukh, J., *et al.* (1998) *J. Infect. Dis.* 178:1193-1197

Cardoso, M., *et al.*, (1998) *J. Med. Virol.* 55:28-34

Davis, G. L., et al., *N. Engl. J. Med.* **339**, 1493-1499.

Farci, P., *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93, 15394-15399

Forns, X., and Bukh, J. (1998) *Viral Hepatitis Rev.* 4, 1-19

Hijikata, M., *et al.* (1991) *Biochem Biophys. Res. Commun.* 175, 220-228

Hoofnagle, J. H. (1997). *Hepatology* **26**, 15S-20S.

Harbersetzer, F., *et al.* (1998) *Virology* 249:32-41

McAllister, J., *et al.* (1998) *J. Virol.* 72: 4893-4905

McHutchison, J. G., et al. (1998) *N. Engl. J. Med.* **339**, 1485-1492.

Puntoriero, G., *et al.* (1998) *EMBO J.* 17:3521-3533

Ray, S.C., *et al.* (1999) *J. Virol.* 73, 2938-2946

Rice, C.M. (1996) in *Fields Virology* (Fields, B.N., Knipe, D.M. and Howley, P.M. eds), pp 931-960, Lippinicot-Raven

Shimizu, Y., *et al.* (1996) *Virology* 223:409-414

Taylor, D., *et al.* (1999) *Science* 285:107-110

Weiner, A.J., *et al.* (1991) *Virology* 180, 842-848

Weiner, A.J., *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89, 3468-3472

Yanagi, M., *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94, 8738-8743

Yanagi, M., *et al.* (1999) *Virology* 262, 250-263

Zibert, A., *et al.* (1997) *Hepatology* 25, 1245-1249

CLAIMS

1. A nucleic acid molecule which encodes a human hepatitis C virus lacking the hypervariable region 1 of HCV envelope 2 protein, said molecule capable of expressing said virus when transfected into cells.

2. The nucleic acid molecule of claim 1, wherein the molecule encodes at least one of the following amino-acids, a histidine at amino acid positions 615, 1143 and 1310, an aspartic acid at position 2875, a methionine at amino acid positions 514 and 1456, and an asparagine at amino acid position 252.

3. A nucleic acid molecule comprising a chimeric virus genome, said genome being a pestivirus or flavivirus genome in which the structural region of the pestivirus or flavivirus has been replaced by the structural region of a hepatitis C virus genome which lacks hypervariable region 1 of HCV envelope 2 protein.

4. The nucleic acid molecule of claim 3, wherein at least one envelope gene from the structural region of the pestivirus or flavivirus genome has been replaced by the E2 gene from the structural region of a hepatitis C virus genome, said HCV E2 gene lacking hypervariable region one.

5. The nucleic acid molecule of claim 3, wherein the structural region of the hepatitis C virus genome encodes at least one of the following amino-acids, a histidine at amino acid position 615, an asparagine at amino acid position 252, and a methionine at position 514.

6. The nucleic acid molecule of claims 3-5, wherein the genome is a pestivirus genome.

7. The nucleic acid molecule of claims 3-5, wherein the pestivirus genome is a bovine viral diarrhea virus genome.

8. The nucleic acid molecule of claims 3-5, wherein the genome is a flavivirus genome.

9. The nucleic acid molecule of claim 8, wherein the flavivirus genome is the genome of a dengue virus.

10. A DNA construct comprising a nucleic acid molecule according to claim 1.

11. A DNA construct comprising a nucleic acid molecule according to claim 3.

12. An RNA transcript of the DNA construct of claim 10.

13. An RNA transcript of the DNA construct of claim 11.

14. A host cell transformed or transfected with the DNA construct of claim 10.

15. A host cell transformed or transfected with the DNA construct of claim 11.

16. A host cell transformed or transfected with RNA transcript of claim 12.

17. A host cell transformed or transfected with RNA transcript of claim 13.

18. A hepatitis C virus envelope 2 polypeptide produced by the cell of claims 14, 15, 16 or 17.

19. A hepatitis C virus produced by the cell of claims 14 or 16.

20. A chimeric virus produced by the cell of claims 15 or 17.

21. A hepatitis C virus whose genome comprises a nucleic acid molecule according to claim 1.

22. A chimeric virus whose genome comprises a nucleic acid molecule according to claim 3.

23. A host cell infected with the virus of claim 21.

24. A host cell infected with the virus of claim 22.

25. An envelope 2 polypeptide encoded by the nucleic acid sequence of claims 1 or 3.

26. Antibody to the polypeptide of claim 25.

27. Antibody to the hepatitis C virus of claim 21.

28. Antibody to the chimeric virus of claim 22.

29. A composition comprising a polypeptide of claim 25 suspended in a suitable amount of a pharmaceutically acceptable diluent or excipient.

30. A composition comprising a virus of claims 21 or 22 suspended in a suitable amount of a pharmaceutically acceptable diluent or excipient.

31. A composition comprising a nucleic acid molecule of claims 1 or 3 suspended in a suitable amount of a pharmaceutically acceptable diluent or excipient.

32. A method of immunizing a mammal, said method comprising immunizing said mammal with the composition of claim 29.

33. A method of immunizing a mammal, said method comprising immunizing said mammal with the composition of claim 30.

34. A method of immunizing a mammal, said method comprising immunizing said mammal with the composition of claim 31.

35. Antibodies produced by the method of claim 32.

36. Antibodies produced by the method of claim 33.

37. Antibodies produced by the method of claim 34.

38. T cells reactive with the virus of claims 21 or 22, said T cells produced by the method of claim 33.

39. T cells reactive with the virus of claims 21 or 22, said T cells produced by the method of claim 34.

40. A chimeric gene comprising, in 5' to 3' order:

(i) an endoplasmic reticulum signal sequence; and

(ii) a coding sequence which encodes a hepatitis C virus (HCV) envelope 2 protein lacking a) hypervariable region 1 of the full-length envelope 2 protein and b) at least the 30 carboxy-terminal amino acids of the full-length envelope 2 protein.

41. The chimeric gene of claim 40, wherein the envelope 2 protein coding sequence lacks about the carboxy terminal 85 amino acids of the full-length envelope 2 protein.

42. The chimeric gene of claim 40, wherein the endoplasmic reticulum signal sequence of i) is amino acids 364-384 of the HCV E1 gene.

43. The chimeric gene of claim 40, wherein the gene further comprises a plasma membrane anchor sequence 3' of the envelope 2 protein coding sequence of ii).

44. The chimeric gene of claim 43, wherein an E1 coding sequence is located between the ER signal sequence of i) and the envelope 2 coding sequence of ii).

45. An expression vector comprising the chimeric gene of claim 40.

46. An expression vector comprising the chimeric gene of claim 43.

47. A host cell transformed or transfected with the expression vector of claim 45.

48. A host cell transformed or transfected with the expression vector of claim 46.

49. A method for expressing a secreted hepatitis C virus envelope 2 protein lacking hypervariable region, said method comprising transforming a host cell with the expression vector of claim 45 under conditions which permit expression of the secreted envelope 2 protein.

50. A method for expressing a hepatitis C virus envelope 2 protein lacking HVR1 on the surface of a cell, said method comprising transforming a host cell with the expression vector of claim 46 under conditions which permit expression of the envelope 2 protein on the cell surface.

51. An envelope 2 protein produced by the method of claim 49.

52. An envelope 2 protein produced by the method of claim 50.

53. A method for immunizing a mammal comprising administering the expression vector of claim 45 to the mammal in an amount effective to stimulate the production of protective antibodies to HCV.

54. A method for immunizing a mammal comprising administering the expression vector of claim 46 to the mammal in an amount effective to stimulate the production of protective antibodies to HCV.

55. A method for immunizing a mammal comprising administering the envelope 2 protein of claim 51 to the mammal in an amount effective to stimulate the production of protective antibodies to HCV.

56. Antibodies produced by the method of claims 53, 54 or 55.

57. A pharmaceutical composition comprising the antibodies of claim 56.

58. A pharmaceutical composition comprising the expression vector of claims 45 or 46.

59. A pharmaceutical composition comprising the protein of claim 51.

60. The nucleic acid molecule of claim 2, wherein the molecule encodes a methionine at amino acid position 514.

FIG. 1A

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCCAGCCCCC	TGATGGGGGC	GACACTCCAC	CATGAATCAC	TCCCCTGTGA	50
GGAACTACTG	TCTTCACGCA	GAAAGCGTCT	AGCCATGGCG	TTAGTATGAG	100
TGTCGTGCAG	CCTCCAGGAC	CCCCCCTCCC	GGGAGAGCCA	TAGTGGTCTG	150
CGGAACCGGT	GAGTACACCG	GAATTGCCAG	GACGACCGGG	TCCTTTCTTG	200
GATAAACCCG	CTCAATGCCT	GGAGATTTGG	GCGTGCCCCC	GCAAGACTGC	250
TAGCCGAGTA	GTGTTGGGTC	GCGAAAGGCC	TTGTGGTACT	GCCTGATAGG	300
GTGCTTGCGA	GTGCCCCCGG	AGGTCTCGTA	GACCGTGCAC	CATGAGCACG	350
AATCCTAAAC	CTCAAAGAAA	AACCAAACGT	AACACCAACC	GTCGCCCACA	400
GGACGTCAAG	TTCCCGGGTG	GCGGTCAGAT	CGTTGGTGGA	GTTTACTTGT	450
TGCCGCGCAG	GGGCCCTAGA	TTGGGTGTGC	GCGCGACGAG	GAAGACTTCC	500
GAGCGGTCGC	AACCTCGAGG	TAGACGTCAG	CCTATCCCCA	AGGCACGTCG	550
GCCCGAGGGC	AGGACCTGGG	CTCAGCCCGG	GTACCCTTGG	CCCCTCTATG	600
GCAATGAGGG	TTGCGGGTGG	GCGGGATGGC	TCCTGTCTCC	CCGTGGCTCT	650
CGGCCTAGCT	GGGGCCCCAC	AGACCCCCGG	CGTAGGTCGC	GCAATTTGGG	700
TAAGGTCATC	GATACCCTTA	CGTGCGGCTT	CGCCGACCTC	ATGGGGTACA	750
TACCGCTCGT	CGGCGCCCCT	CTTGGAGGCG	CTGCCAGGGC	CCTGGCGCAT	800
GGCGTCCGGG	TTCTGGAAGA	CGGCGTGAAC	TATGCAACAG	GGAACCTTCC	850
TGGTTGCTCT	TTCTCTATCT	TCCTTCTGGC	CCTGCTCTCT	TGCCTGACTG	900
TGCCCGCTTC	AGCCTACCAA	GTGCGCAATT	CCTCGGGGCT	TTACCATGTC	950
ACCAATGATT	GCCCTAACTC	GAGTATTGTG	TACGAGGCGG	CCGATGCCAT	1000
CCTGCACACT	CCGGGGTGTG	TCCCTTGCGT	TCGCGAGGGT	AACGCCTCGA	1050
GGTGTTGGGT	GGCGGTGACC	CCCACGGTGG	CCACCAGGGA	CGGCAAACTC	1100
CCCACAACGC	AGCTTCGACG	TCATATCGAT	CTGCTTGTCG	GGAGCGCCAC	1150
CCTCTGCTCG	GCCCTCTACG	TGGGGGACCT	GTGCGGGTCT	GTCTTTCTTG	1200
TTGGTCAACT	GTTTACCTTC	TCTCCCAGGC	GCCACTGGAC	GACGCAAGAC	1250
TGCAATTGTT	CTATCTATCC	CGGCCATATA	ACGGGTCATC	GCATGGCATG	1300
GGATATGATG	ATGAACTGGT	CCCCTACGGC	AGCGTTGGTG	GTAGCTCAGC	1350
TGCTCCGGAT	CCCACAAGCC	ATCATGGACA	TGATCGCTGG	TGCTCACTGG	1400
GGAGTCCTGG	CGGGCATAGC	GTATTTCTCC	ATGGTGGGGA	ACTGGGCGAA	1450
GGTCCTGGTA	GTGCTGCTGC	TATTTGCCGG	CGTCGACGCG	ATCCAACTGA	1500
TCAACACCAA	CGGCAGTTGG	CACATCAATA	GCACGGCCTT	GAATTGCAAT	1550
GAAAGCCTTA	ACACCGGCTG	GTTAGCAGGG	CTCTTCTATC	AACACAAATT	1600
CAACTCTTCA	GGCTGTCCTG	AGAGGTTGGC	CAGCTGCCGA	CGCCTTACCG	1650
ATTTTGCCCA	GGGCTGGGGT	CCTATCAGTT	ATGCCAACGG	AAGCGGCCTC	1700
GACGAACGCC	CCTACTGCTG	GCACTACCCT	CCAAGACCTT	GTGGCATTGT	1750
GCCCGCAAAG	AGCGTGTGTG	GCCCGGTATA	TTGCTTCACT	CCCAGCCCCG	1800

FIG. 1B

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGGTGGTGGG	AACGACCGAC	AGGTCGGGCG	CGCCTACCTA	CAGCTGGGGT	1850
GCAAATGATA	CGGATGTCTT	CGTCCTTAAC	AACACCAGGC	CACCGCTGGG	1900
CAATTGGTTC	GGTTGTACCT	GGATGAACTC	AACTGGATTC	ACCAAAGTGT	1950
GCGGAGCGCC	CCCTTGTGTC	ATCGGAGGGG	TGGGCAACAA	CACCTTGCTC	2000
TGCCCCACTG	ATTGCTTCCG	CAAACATCCG	GAAGCCACAT	ACTCTCGGTG	2050
CGGCTCCGGT	CCCTGGATTA	CACCCAGGTG	CATGGTCGAC	TACCCGTATA	2100
GGCTTTGGCA	CTATCCTTGT	ACCATCAATT	ACACCATATT	CAAAGTCAGG	2150
ATGTACGTGG	GAGGGGTCGA	GCACAGGCTG	GAAGCGGCCT	GCAACTGGAC	2200
GCGGGGCGAA	CGCTGTGATC	TGGAAGACAG	GGACAGGTCC	GAGCTCAGCC	2250
CGTTGCTGCT	GTCCACCACA	CAGTGGCAGG	TCCTTCCGTG	TTCTTTCACG	2300
ACCCTGCCAG	CCTTGTCCAC	CGGCCTCATC	CACCTCCACC	AGAACATTGT	2350
GGACGTGCAG	TACTTGTACG	GGGTAGGGTC	AAGCATCGCG	TCCTGGGCCA	2400
TTAAGTGGGA	GTACGTCGTT	CTCCTGTTCC	TTCTGCTTGC	AGACGCGCGC	2450
GTCTGCTCCT	GCTTGTGGAT	GATGTTACTC	ATATCCCAAG	CGGAGCCGGC	2500
TTTGGAGAAC	CTCGTAATAC	TCAATGCAGC	ATCCCTGGCC	GGGACGCACG	2550
GTCTTGTGTC	CTTCCTCGTG	TTCTTCTGCT	TTGCGTGGTA	TCTGAAGGGT	2600
AGGTGGGTGC	CCGGAGCGGT	CTACGCCCTC	TACGGGATGT	GGCCTCTCCT	2650
CCTGCTCCTG	CTGGCGTTGC	CTCAGCGGGC	ATACGCACTG	GACACGGAGG	2700
TGGCCGCGTC	GTGTGGCGGC	GTTGTTCTTG	TCGGGTTAAT	GGCGCTGACT	2750
CTGTCGCCAT	ATTACAAGCG	CTATATCAGC	TGGTGCATGT	GGTGGCTTCA	2800
GTATTTTCTG	ACCAGAGTAG	AAGCGCAACT	GCACGTGTGG	GTTCCCCCCC	2850
TCAACGTCCG	GGGGGGCGGC	GATGCCGTCA	TCTTACTCAT	GTGTGTAGTA	2900
CACCCGACCC	TGGTATTTGA	CATCACCAAA	CTACTCCTGG	CCATCTTCGG	2950
ACCCCTTTGG	ATTCTTCAAG	CCAGTTTGCT	TAAAGTCCCC	TACTTCGTGC	3000
GCGTTCAAGG	CCTTCTCCGG	ATCTGCGCGC	TAGCGCGGAA	GATAGCCGGA	3050
GGTCATTACG	TGCAAATGGC	CATCATCAAG	TTAGGGCGGC	TTACTGGCAC	3100
CTATGTGTAT	AACCATCTCA	CCCCTCTTCG	AGACTGGGCG	CACAACGGCC	3150
TGCGAGATCT	GGCCGTGGCT	GTGGAACCAG	TCGTCTTCTC	CCGAATGGAG	3200
ACCAAGCTCA	TCACGTGGGG	GGCAGATACC	GCCGCGTGCG	GTGACATCAT	3250
CAACGGCTTG	CCCGTCTCTG	CCCGTAGGGG	CCAGGAGATA	CTGCTTGGGC	3300
CAGCCGACGG	AATGGTCTCC	AAGGGGTGGA	GGTTGCTGGC	GCCCATCACG	3350
GCGTACGCCC	AGCAGACGAG	AGGCCTCCTA	GGGTGTATAA	TCACCAGCCT	3400
GACTGGCCGG	GACAAAAACC	AAGTGGAGGG	TGAGGTCCAG	ATCGTGTCAA	3450
CTGCTACCCA	AACCTTCCTG	GCAACGTGCA	TCAATGGGGT	ATGCTGGACT	3500
GTCTACCACG	GGGCCGGAAC	GAGGACCATC	GCATCACCCA	AGGGTCCTGT	3550
CATCCAGATG	TATACCAATG	TGGACCAAGA	CCTTGTGGGC	TGGCCCGCTC	3600

FIG. 1C

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CTCAAGGTTC	CCGCTCATTG	ACACCCTGTA	CCTGCGGCTC	CTCGGACCTT	3650
TACCTGGTCA	CGAGGCACGC	CGATGTCATT	CCCGTGCGCC	GGCGAGGTGA	3700
TAGCAGGGGT	AGCCTGCTTT	CGCCCCGGCC	CATTTCCTAC	TTGAAAGGCT	3750
CCTCGGGGGG	TCCGCTGTTG	TGCCCCGCGG	GACACGCCGT	GGGCCTATTC	3800
AGGGCCGCGG	TGTGCACCCG	TGGAGTGGCT	AAAGCGGTGG	ACTTTATCCC	3850
TGTGGAGAAC	CTAGGGACAA	CCATGAGATC	CCCGGTGTTC	ACGGACAACT	3900
CCTCTCCACC	AGCAGTGCCC	CAGAGCTTCC	AGGTGGCCCA	CCTGCATGCT	3950
CCCACCGGCA	GCGGTAAGAG	CACCAAGGTC	CCGGCTGCGT	ACGCAGCCCA	4000
GGGCTACAAG	GTGTTGGTGC	TCAACCCCTC	TGTTGCTGCA	ACGCTGGGCT	4050
TTGGTGCTTA	CATGTCCAAG	GCCCATGGGG	TTGATCCTAA	TATCAGGACC	4100
GGGGTGAGAA	CAATTACCAC	TGGCAGCCCC	ATCACGTACT	CCACCTACGG	4150
CAAGTTCCTT	GCCGACGGCG	GGTGCTCAGG	AGGTGCTTAT	GACATAATAA	4200
TTTGTGACGA	GTGCCACTCC	ACGGATGCCA	CATCCATCTT	GGGCATCGGC	4250
ACTGTCCTTG	ACCAAGCAGA	GACTGCGGGG	GCGAGACTGG	TTGTGCTCGC	4300
CACTGCTACC	CCTCCGGGCT	CCGTCACTGT	GTCCCATCCT	AACATCGAGG	4350
AGGTTGCTCT	GTCCACCACC	GGAGAGATCC	CCTTTTACGG	CAAGGCTATC	4400
CCCCTCGAGG	TGATCAAGGG	GGGAAGACAT	CTCATCTTCT	GCCACTCAAA	4450
GAAGAAGTGC	GACGAGCTCG	CCGCGAAGCT	GGTCGCATTG	GGCATCAATG	4500
CCGTGGCCTA	CTACCGCGGT	CTTGACGTGT	CTGTCATCCC	GACCAGCGGC	4550
GATGTTGTCG	TCGTGTCGAC	CGATGCTCTC	ATGACTGGCT	TTACCGGCGA	4600
CTTCGACTCT	GTGATAGACT	GCAACACGTG	TGTCACTCAG	ACAGTCGATT	4650
TCAGCCTTGA	CCCTACCTTT	ACCATTGAGA	CAACCACGCT	CCCCCAGGAT	4700
GCTGTCTCCA	GGACTCAACG	CCGGGGCAGG	ACTGGCAGGG	GGAAGCCAGG	4750
CATCTATAGA	TTTGTGGCAC	CGGGGGAGCG	CCCCTCCGGC	ATGTTCGACT	4800
CGTCCGTCCT	CTGTGAGTGC	TATGACGCGG	GCTGTGCTTG	GTATGAGCTC	4850
ACGCCCGCCG	AGACTACAGT	TAGGCTACGA	GCGTACATGA	ACACCCCCGG	4900
GCTTCCCGTG	TGCCAGGACC	ATCTTGAATT	TTGGGAGGGC	GTCTTTACGG	4950
GCCTCACTCA	TATAGATGCC	CACTTTTTAT	CCCAGACAAA	GCAGAGTGGG	5000
GAGAACTTTC	CTTACCTGGT	AGCGTACCAA	GCCACCGTGT	GCGCTAGGGC	5050
TCAAGCCCCT	CCCCCATCGT	GGGACCAGAT	GTGGAAGTGT	TTGATCCGCC	5100
TTAAACCCAC	CCTCCATGGG	CCAACACCCC	TGCTATACAG	ACTGGGCGCT	5150
GTTCAGAATG	AAGTCACCCT	GACGCACCCA	ATCACCAAAT	ACATCATGAC	5200
ATGCATGTCG	GCCGACCTGG	AGGTCGTCAC	GAGCACCTGG	GTGCTCGTTG	5250
GCGGCGTCCT	GGCTGCTCTG	GCCGCGTATT	GCCTGTCAAC	AGGCTGCGTG	5300
GTCATAGTGG	GCAGGATCGT	CTTGTCCGGG	AAGCCGGCAA	TTATACCTGA	5350
CAGGGAGGTT	CTCTACCAGG	AGTTCGATGA	GATGGAAGAG	TGCTCTCAGC	5400

FIG. 1D

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ACTTACCGTA	CATCGAGCAA	GGGATGATGC	TCGCTGAGCA	GTTCAAGCAG	5450
AAGGCCCTCG	GCCTCCTGCA	GACCGCGGTCC	CGCCATGCAG	AGGTTATCAC	5500
CCCTGCTGTC	CAGACCAACT	GGCAGAAACT	CGAGGTCTTT	TGGGCGAAGC	5550
ACATGTGGAA	TTTCATCAGT	GGGATACAAT	ACTTGGCGGG	CCTGTCAACG	5600
CTGCCTGGTA	ACCCCGCCAT	TGCTTCATTG	ATGGCTTTTA	CAGCTGCCGT	5650
CACCAGCCCA	CTAACCACTG	GCCAAACCCT	CCTCTTCAAC	ATATTGGGGG	5700
GGTGGGTGGC	TGCCCAGCTC	GCCGCCCCCG	GTGCCGCTAC	TGCCTTTGTG	5750
GGTGCTGGCC	TAGCTGGCGC	CGCCATCGGC	AGCGTTGGAC	TGGGGAAGGT	5800
CCTCGTGGAC	ATTCTTGCAG	GGTATGGCGC	GGGCGTGGCG	GGAGCTCTTG	5850
TAGCATTCAA	GATCATGAGC	GGTGAGGTCC	CCTCCACGGA	GGACCTGGTC	5900
AATCTGCTGC	CCGCCATCCT	CTCGCCTGGA	GCCCTTGTAG	TCGGTGTGGT	5950
CTGCGCAGCA	ATACTGCGCC	GGCACGTTGG	CCCGGGCGAG	GGGGCAGTGC	6000
AATGGATGAA	CCGGCTAATA	GCCTTCGCCT	CCCGGGGGAA	CCATGTTTCC	6050
CCCACGCACT	ACGTGCCGGA	GAGCGATGCA	GCCGCCCGCG	TCACTGCCAT	6100
ACTCAGCAGC	CTCACTGTAA	CCCAGCTCCT	GAGGCGACTG	CATCAGTGGA	6150
TAAGCTCGGA	GTGTACCACT	CCATGCTCCG	GTTCCTGGCT	AAGGGACATC	6200
TGGGACTGGA	TATGCGAGGT	GCTGAGCGAC	TTTAAGACCT	GGCTGAAAGC	6250
CAAGCTCATG	CCACAACTGC	CTGGGATTCC	CTTTGTGTCC	TGCCAGCGCG	6300
GGTATAGGGG	GGTCTGGCGA	GGAGACGGCA	TTATGCACAC	TCGCTGCCAC	6350
TGTGGAGCTG	AGATCACTGG	ACATGTCAAA	AACGGGACGA	TGAGGATCGT	6400
CGGTCCTAGG	ACCTGCAGGA	ACATGTGGAG	TGGGACGTTC	CCCATTAACG	6450
CCTACACCAC	GGGCCCCTGT	ACTCCCCTTC	CTGCGCCGAA	CTATAAGTTC	6500
GCGCTGTGGA	GGGTGTCTGC	AGAGGAATAC	GTGGAGATAA	GGCGGGTGGG	6550
GGACTTCCAC	TACGTATCGG	GTATGACTAC	TGACAATCTT	AAATGCCCGT	6600
GCCAGATCCC	ATCGCCCGAA	TTTTTCACAG	AATTGGACGG	GGTGCGCCTA	6650
CACAGGTTTG	CGCCCCCTTG	CAAGCCCTTG	CTGCGGGAGG	AGGTATCATT	6700
CAGAGTAGGA	CTCCACGAGT	ACCCGGTGGG	GTCGCAATTA	CCTTGCGAGC	6750
CCGAACCGGA	CGTAGCCGTG	TTGACGTCCA	TGCTCACTGA	TCCCTCCCAT	6800
ATAACAGCAG	AGGCGGCCGG	GAGAAGGTTG	GCGAGAGGGT	CACCCCCTTC	6850
TATGGCCAGC	TCCTCGGCTA	GCCAGCTGTC	CGCTCCATCT	CTCAAGGCAA	6900
CTTGCACCGC	CAACCATGAC	TCCCCTGACG	CCGAGCTCAT	AGAGGCTAAC	6950
CTCCTGTGGA	GGCAGGAGAT	GGGCGGCAAC	ATCACCAGGG	TTGAGTCAGA	7000
GAACAAAGTG	GTGATTCTGG	ACTCCTTCGA	TCCGCTTGTG	GCAGAGGAGG	7050
ATGAGCGGGA	GGTCTCCGTA	CCTGCAGAAA	TTCTGCGGAA	GTCTCGGAGA	7100
TTCGCCCGGG	CCCTGCCCGT	CTGGGCGCGG	CCGGACTACA	ACCCCCCGCT	7150
AGTAGAGACG	TGGAAAAAGC	CTGACTACGA	ACCACCTGTG	GTCCATGGCT	7200

FIG. 1E

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCCCGCTACC	ACCTCCACGG	TCCCCTCCTG	TGCCTCCGCC	TCGGAAAAAG	7250
CGTACGGTGG	TCCTCACCGA	ATCAACCCTA	TCTACTGCCT	TGGCCGAGCT	7300
TGCCACCAAA	AGTTTTGGCA	GCTCCTCAAC	TTCCGGCATT	ACGGGCGACA	7350
ATACGACAAC	ATCCTCTGAG	CCCGCCCCTT	CTGGCTGCCC	CCCCGACTCC	7400
GACGTTGAGT	CCTATTCTTC	CATGCCCCCC	CTGGAGGGGG	AGCCTGGGGA	7450
TCCGGATCTC	AGCGACGGGT	CATGGTCGAC	GGTCAGTAGT	GGGGCCGACA	7500
CGGAAGATGT	CGTGTGCTGC	TCAATGTCTT	ATTCCTGGAC	AGGCGCACTC	7550
GTCACCCCGT	GCGCTGCGGA	AGAACAAAAA	CTGCCCATCA	ACGCACTGAG	7600
CAACTCGTTG	CTACGCCATC	ACAATCTGGT	GTATTCCACC	ACTTCACGCA	7650
GTGCTTGCCA	AAGGCAGAAG	AAAGTCACAT	TTGACAGACT	GCAAGTTCTG	7700
GACAGCCATT	ACCAGGACGT	GCTCAAGGAG	GTCAAAGCAG	CGGCGTCAAA	7750
AGTGAAGGCT	AACTTGCTAT	CCGTAGAGGA	AGCTTGCAGC	CTGACGCCCC	7800
CACATTCAGC	CAAATCCAAG	TTTGGCTATG	GGGCAAAAGA	CGTCCGTTGC	7850
CATGCCAGAA	AGGCCGTAGC	CCACATCAAC	TCCGTGTGGA	AAGACCTTCT	7900
GGAAGACAGT	GTAACACCAA	TAGACACTAC	CATCATGGCC	AAGAACGAGG	7950
TTTTCTGCGT	TCAGCCTGAG	AAGGGGGGTC	GTAAGCCAGC	TCGTCTCATC	8000
GTGTTCCCCG	ACCTGGGCGT	GCGCGTGTGC	GAGAAGATGG	CCCTGTACGA	8050
CGTGGTTAGC	AAGCTCCCCC	TGGCCGTGAT	GGGAAGCTCC	TACCGATTCC	8100
AATACTCACC	AGGACAGCGG	GTTGAATTCC	TCGTGCAAGC	GTGGAAGTCC	8150
AAGAAGACCC	CGATGGGGTT	CTCGTATGAT	ACCCGCTGTT	TTGACTCCAC	8200
AGTCACTGAG	AGCGACATCC	GTACGGAGGA	GGCAATTTAC	CAATGTTGTG	8250
ACCTGGACCC	CCAAGCCCGC	GTGGCCATCA	AGTCCCTCAC	TGAGAGGCTT	8300
TATGTTGGGG	GCCCTCTTAC	CAATTCAAGG	GGGGAAAACT	GCGGCTACCG	8350
CAGGTGCCGC	GCGAGCGGCG	TACTGACAAC	TAGCTGTGGT	AACACCCTCA	8400
CTTGCTACAT	CAAGGCCCGG	GCAGCCTGTC	GAGCCGCAGG	GCTCCAGGAC	8450
TGCACCATGC	TCGTGTGTGG	CGACGACTTA	GTCGTTATCT	GTGAAAGTGC	8500
GGGGGTCCAG	GAGGACGCGG	CGAGCCTGAG	AGCCTTCACG	GAGGCTATGA	8550
CCAGGTACTC	CGCCCCCCCC	GGGGACCCCC	CACAACCAGA	ATACGACTTG	8600
GAGCTTATAA	CATCATGCTC	CTCCAACGTG	TCAGTCGCCC	ACGACGGCGC	8650
TGGAAAGAGG	GTCTACTACC	TTACCCGTGA	CCCTACAACC	CCCCTCGCGA	8700
GAGCCGCGTG	GGAGACAGCA	AGACACACTC	CAGTCAATTC	CTGGCTAGGC	8750
AACATAATCA	TGTTTGCCCC	CACACTGTGG	GCGAGGATGA	TACTGATGAC	8800
CCATTTCTTT	AGCGTCCTCA	TAGCCAGGGA	TCAGCTTGAA	CAGGCTCTTA	8850
ACTGTGAGAT	CTACGGAGCC	TGCTACTCCA	TAGAACCACT	GGATCTACCT	8900
CCAATCATTC	AAAGACTCCA	TGGCCTCAGC	GCATTTTCAC	TCCACAGTTA	8950
CTCTCCAGGT	GAAATCAATA	GGGTGGCCGC	ATGCCTCAGA	AAACTTGGGG	9000

FIG. 1F

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCCCGCCCTT	GCGAGCTTGG	AGACACCGGG	CCCGGAGCGT	CCGCGCTAGG	9050
CTTCTGTCCA	GAGGAGGCAG	GGCTGCCATA	TGTGGCAAGT	ACCTCTTCAA	9100
CTGGGCAGTA	AGAACAAAGC	TCAAACTCAC	TCCAATAGCG	GCCGCTGGCC	9150
GGCTGGACTT	GTCCGGTTGG	TTCACGGCTG	GCTACAGCGG	GGGAGACATT	9200
TATCACAGCG	TGTCTCATGC	CCGGCCCCGC	TGGTTCTGGT	TTTGCCTACT	9250
CCTGCTCGCT	GCAGGGGTAG	GCATCTACCT	CCTCCCCAAC	CGATGAAGGT	9300
TGGGGTAAAC	ACTCCGGCCT	CTTAAGCCAT	TTCCTGTTTT	TTTTTTTTTT	9350
TTTTTTTTTT	TTTTCTTTTT	TTTTTTCTTT	CCTTTCCTTC	TTTTTTTCCT	9400
TTCTTTTTCC	CTTCTTTAAT	GGTGGCTCCA	TCTTAGCCCT	AGTCACGGCT	9450
AGCTGTGAAA	GGTCCGTGAG	CCGCATGACT	GCAGAGAGTG	CTGATACTGG	9500
CCTCTCTGCA	GATCATGT				9518

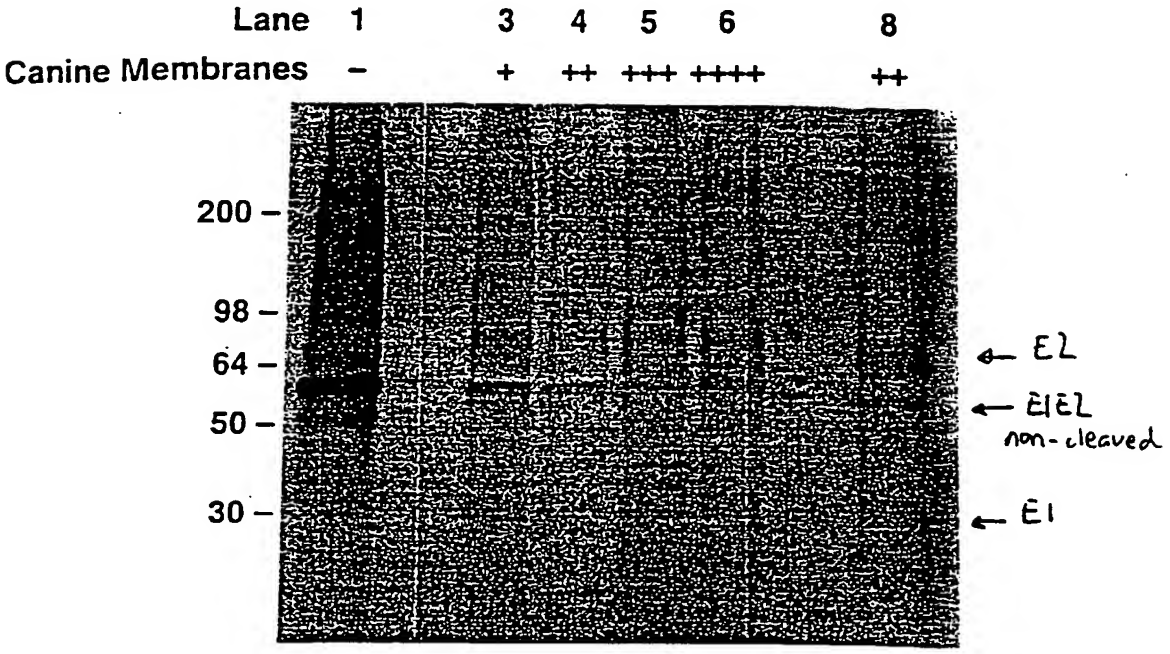
FIG. 1G

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
MSINPKPQRK	TKRNINRRPQ	DVKFPGGGQI	VGGVYLLPRR	GPRLGVRATR	50
KTSERSQPRG	RRQPIPKARR	PEGRIWAQPG	YPWPLYGNEG	CGWAGWLLSP	100
RGSRPSWGPT	DPRRRSRNLG	KVIDTLTCGF	ADLMGYIPLV	GAPLGGAARA	150
LAHGVRVLED	GVNYATGNLP	GCSFSIFLLA	LLSCLTVPAS	AYQVRNSSGL	200
YHVINDCPNS	SIVYEAADAI	LHTPGCVPCV	REGNASRCWV	AVTPTVATRD	250
GKLPTTQLRR	HIDLLVGSAT	LCSALYVGDL	CGSVFLVGQL	FTFSPRRHWT	300
TQDCNCSIYP	GHITGHRMAW	DMMMNWSPTA	ALVVAQLLRI	PQAIMDMIAG	350
AHWGVLAGIA	YFSMVGNWAK	VLVVLLLFAG	VDAIQLTNTN	GSWHINSTAL	400
NCNESLNTGW	LAGLFYQHKF	NSSGCPERLA	SCRRLTDFAQ	GWGPISYANG	450
SGLDERPYCW	HYPPRPCGIV	PAKSVCGPVY	CFTPSPVVVG	TTDRSGAPTY	500
SWGANDTDVF	VLNNTRPPLG	NWFGCTWMNS	TGFTKVCGAP	PCVIGGVGNN	550
TLLCPTDCFR	KHPEATYSRC	GSGPWITPRC	MVDYPYRLWH	YPCTINYTIF	600
KVRMYVGGVE	HRLEAACNWT	RGERCDLEDR	DRSELSPLLL	STTQWQVLPC	650
SFTTLPALST	GLIHLHQNIV	DVQYLYGVGS	SIASWAIKWE	YVVLLFLLLA	700
DARVCSCLWM	MLLISQAEAA	LENLVILNAA	SLAGTHGLVS	FLVFFCFAWY	750
LKGRWVPGAV	YALYGMWPLL	LLLLALPQRA	YALDTEVAAS	CGGVVLVGLM	800
ALTLSPYYKR	YISWCMWWLQ	YFLTRVEAQL	HVWVPPLNVR	GGRDAVILLM	850
CVVHPTLVFD	ITKLLLAIFG	PLWILQASLL	KVPYFVRVQG	LLRICALARK	900
IAGGHYVQMA	IIKLGALTGT	YVYNHLTPLR	DWAHNGLRDL	AVAVEPVVFS	950
RMEIKLITWG	ADIAACGDII	NGLPVSARRG	QEILLGPADG	MVSKGWRLLA	1000
PITAYAQQTR	GLLGCIITSL	TGRDKNQVEG	EVQIVSTATQ	TFLATCINGV	1050
CWTVYHGAGT	RTIASPKGPV	IQMYTNVDQD	LVGWPAPQGS	RSLTPCTCGS	1100
SDLYLVTRHA	DVIPVRRRGD	SRGSLLSPRP	ISYLKGSSGG	PLLCPAGHAV	1150
GLFRAAVCTR	GVAKAVDFIP	VENLGTTMRS	PVFTDNSSPP	AVPQSFQVAH	1200
LHAPIGSGKS	TKVPAAYAAQ	GYKVLVLNPS	VAATLGFGAY	MSKAHGVDPN	1250
IRTGVRTITT	GSPITYSTYG	KFLADGGCSG	GAYDIIICDE	CHSTDATSIL	1300
GIGTVLDQAE	TAGARLVVLA	TATPPGSVTV	SHPNIEEVAL	STTGEIPFYG	1350
KAIPLEVIKG	GRHLIFCHSK	KKCDELAAKL	VALGINAVAY	YRGLDVSVIP	1400
TSGDVVVVST	DALMTGFTGD	FDSVIDCNTC	VTQTVDFSLD	PTFTIETTTL	1450
PQDAVSRTQR	RGRTGRGKPG	IYRFVAPGER	PSGMFDSSVL	CECYDAGCAW	1500
YELTPAETTV	RLRAYMNTPG	LPVCQDHLEF	WEGVFTGLTH	IDAHFLSQTK	1550
QSGENFPYLV	AYQATVCARA	QAPPPSWDQM	WKCLIRLKPT	LHGPTPLLYR	1600
LGAVQNEVTL	THPITKYIMT	CMSADLEVVT	STWVLVGGVL	AALAAYCLST	1650
GCVVIVGRIV	LSGKPAIIPD	REVLYQEFDE	MEECSQHLPY	IEQGMMLAEQ	1700
FKQKALGLLQ	TASRHAEVIT	PAVQTNWQKL	EVFWAKHMWN	FISGIQYLAG	1750
LSTLPGNPAI	ASLMAFTAAV	TSPLTTGQTL	LFNILGGWVA	AQLAAPGAAT	1800
AFVGAGLAGA	AIGSVGLGKV	LVDILAGYGA	GVAGALVAFK	IMSGEVPSTE	1850
DLVNLLPAIL	SPGALVVGVV	CAAILRRHVG	PGEGAVQWMN	RLIAFASRGN	1900

FIG. 1H

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
HVSPTHYVPE	SDAAARVTAI	LSSLTVTQLL	RRLHQWISSE	CTTPCSGSWL	1950
RDIWDWICEV	LSDFKIWLKA	KLMPQLPGIP	FVSCQRGYRG	VWRGDGIMHT	2000
RCHCGAEITG	HVKNGTMRIV	GPRTCRNMWS	GTFPINAYTT	GPCTPLPAPN	2050
YKFALWRVSA	EEYVEIRRVG	DFHYVSGMTT	DNLKCPCQIP	SPEFFTELDG	2100
VRLHRFAPPC	KPLLREEVSF	RVGLHEYPVG	SQLPCEPEPD	VAVLTSMLTD	2150
PSHITAEAAG	RRLARGSPPS	MASSSASQLS	APSLKATCTA	NHDSPDAELI	2200
EANLLWRQEM	GGNITRVESE	NKVVILDSFD	PLVAEEDERE	VSVPAEILRK	2250
SRRFARALPV	WARPDYNPPL	VETWKKPDYE	PPVVHGCPLP	PPRSPPVPPP	2300
RKKRTVVLTE	STLSTALAEL	ATKSFGSSST	SGITGDNTTT	SSEPAPSGCP	2350
PDSDVESYSS	MPPLEGEPGD	PDLSDGSWST	VSSGADTEDV	VCCSMSYSWT	2400
GALVTPCAAE	EQKLPINALS	NSLLRHHNLV	YSTTSRSACQ	RQKKVTFDRL	2450
QVLDSHYQDV	LKEVKAAASK	VKANLLSVEE	ACSLTPPHSA	KSKFGYGAKD	2500
VRCHARKAVA	HINSVWKDLL	EDSVTPIDTT	IMAKNEVFCV	QPEKGGRKPA	2550
RLIVFPDLGV	RVCEKMALYD	VVSKLPLAVM	GSSYGFQYSP	GQRVEFLVQA	2600
WKSKKTPMGF	SYDTRCFDST	VTESDIRTEE	AIYQCCDLDP	QARVAIKSLT	2650
ERLYVGGPLT	NSRGENCGYR	RCRASGVLTT	SCGNTLTCYI	KARAACRAAG	2700
LQDCTMLVCG	DDLVVICESA	GVQEDAASLR	AFTEAMTRYS	APPGDPPQPE	2750
YDLELITSCS	SNVSVAHDGA	GKRVYYLTRD	PTTPLARAAW	ETARHTPVNS	2800
WLGNIIMFAP	TLWARMILMT	HFFSVLIARD	QLEQALNCEI	YGACYSIEPL	2850
DLPPIIQRLH	GLSAFSLHSY	SPGEINRVAA	CLRKLGVPPL	RAWRHRARSV	2900
RARLLSRGGR	AAICGKYLFN	WAVRTKLKLT	PIAAAGRLDL	SGWFTAGYSG	2950
GDIYHSVSHA	RPRWFWFCLL	LLAAGVGIYL	LPNR		2984

FIG. 2


In Vitro Transcription Translation
Lane 1 3 4 5 6 8
Canine Membranes – + ++ +++ ++++ ++
200 –
98 –
64 –
50 –
30 –
← E2
← E1E2
non-cleaved
← E1

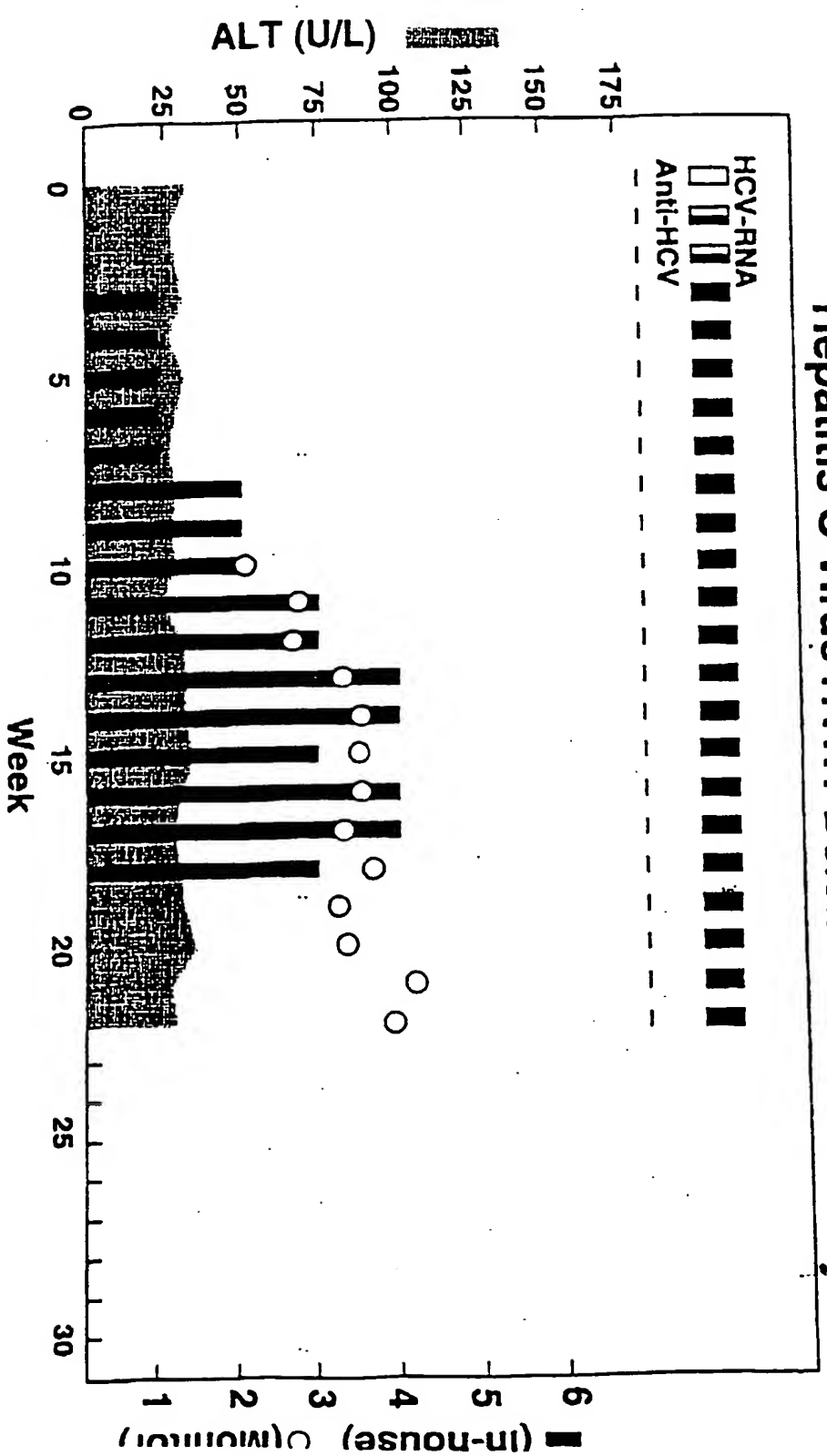
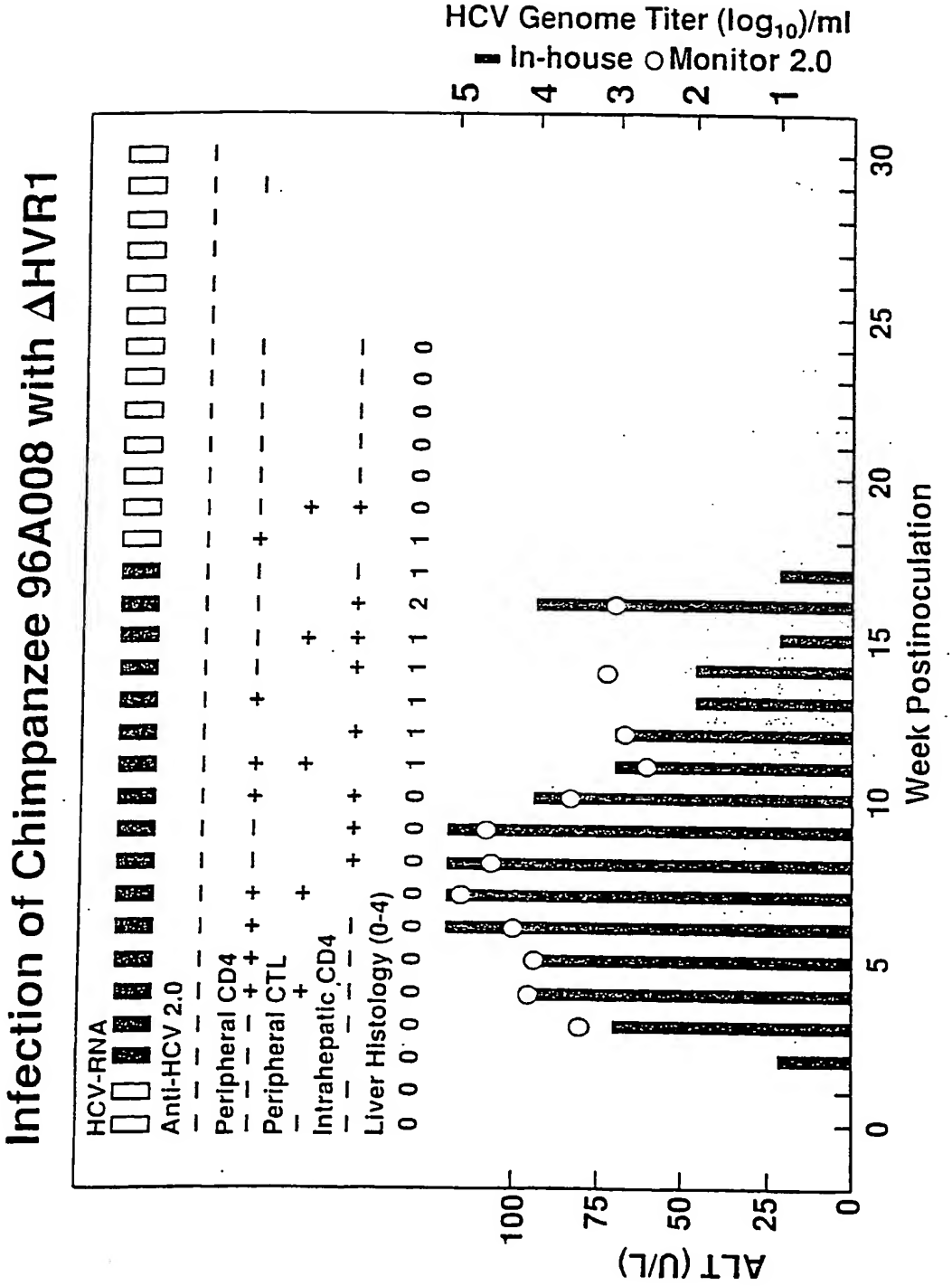

Chimpanzee 1590: Transfection with RNA Transcripts of
Hepatitis C Virus HVR1 Deletion Mutant
ALT (U/L)
HCV-RNA
Anti-HCV
Week
0
5
10
15
20
25
30
0
25
50
75
100
125
150
175
1
2
3
4
5
6

FIG. 3

Figure 4


Infection of Chimpanzee 96A008 with ΔHVR1
HCV Genome Titer (log_{10})/ml
In-house
Monitor 2.0
HCV-RNA
Anti-HCV 2.0
Peripheral CD4
Peripheral CTL
Intrahepatic CD4
Liver Histology (0-4)
ALT (U/L)
Week Postinoculation

SEQUENCE LISTING

<110> FORNS, XAVIER
BUKH, JENS
EMERSON, SUZANNE U.
PURCELL, ROBERT H.

<120> NUCLEIC ACID MOLECULES ENCODING HEPATITIS C VIRUS WHICH LACK HYPERVARIABLE REGION ONE OF E2 AND USES THEREOF

<130> 2026-431PC

<140> To Be Assigned
<141> 2000-09-22

<150> 60/155,823
<151> 1999-09-23

<160> 12

<170> PatentIn Ver. 2.1

<210> 1
<211> 9518
<212> DNA
<213> Hepatitis C virus

<400> 1
gccagccccc tgatgggggc gacactccac catgaatcac tcccctgtga ggaactactg 60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgtcgtgcag cctccaggac 120
cccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg tcctttcttg gataaacccg ctcaatgcct ggagatttgg gcgtgccccc 240
gcaagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaacgt aacaccaacc gtcgcccaca ggacgtcaag ttcccgggtg 420
gcggtcagat cgttggtgga gtttacttgt tgccgcgcag gggccctaga ttgggtgtgc 480
gcgcgacgag gaagacttcc gagcggtcgc aacctcgagg tagacgtcag cctatcccca 540
aggcacgtcg gcccgagggc aggacctggg ctcagcccgg gtaccccttgg cccctctatg 600
gcaatgaggg ttgcgggtgg gcgggatggc tcctgtctcc ccgtggctct cggcctagct 660
ggggccccac agacccccgg cgtaggtcgc gcaatttggg taaggtcatc gataccctta 720
cgtgcggctt cgccgacctc atggggtaca taccgctcgt cggcgccccct cttggaggcg 780
ctgccagggc cctggcgcat ggcgtccggg ttctggaaga cggcgtgaac tatgcaacag 840
ggaaccttcc tggttgctct ttctctatct tccttctggc cctgctctct tgcctgactg 900
tgcccgcttc agcctaccaa gtgcgcaatt cctcggggct ttaccatgtc accaatgatt 960
gccctaactc gagtattgtg tacgaggcgg ccgatgccat cctgcacact ccggggtgtg 1020
tcccttgcgt tcgcgagggt aacgcctcga ggtgttgggt ggcggtgacc cccacggtgg 1080
ccaccaggga cggcaaactc cccacaacgc agcttcgacg tcatatcgat ctgcttgtcg 1140
ggagcgccac cctctgctcg gccctctacg tgggggacct gtgcgggtct gtctttcttg 1200

ttggtcaact gtttaccttc tctcccaggc gccactggac gacgcaagac tgcaattgtt 1260
ctatctatcc cggccatata acgggtcatc gcatggcatg ggatatgatg atgaactggt 1320
cccctacggc agcgttggtg gtagctcagc tgctccggat cccacaagcc atcatggaca 1380
tgatcgctgg tgctcactgg ggagtcctgg cgggcatagc gtatttctcc atggtgggga 1440
actgggcgaa ggtcctggta gtgctgctgc tatttgccgg cgtcgacgcg atccaactga 1500
tcaacaccaa cggcagttgg cacatcaata gcacggcctt gaattgcaat gaaagcctta 1560
acaccggctg gttagcaggg ctcttctatc aacacaaatt caactcttca ggctgtcctg 1620
agaggttggc cagctgccga cgccttaccg attttgccca gggctggggt cctatcagtt 1680
atgccaacgg aagcggcctc gacgaacgcc cctactgctg gcactaccct ccaagacctt 1740
gtggcattgt gcccgcaaag agcgtgtgtg gcccggtata ttgcttcact cccagccccg 1800
tggtggtggg aacgaccgac aggtcgggcg cgcctaccta cagctggggt gcaaatgata 1860
cggatgtctt cgtccttaac aacaccaggc caccgctggg caattggttc ggttgtacct 1920
ggatgaactc aactggattc accaaagtgt gcggagcgcc cccttgtgtc atcggagggg 1980
tgggcaacaa caccttgctc tgccccactg attgcttccg caaacatccg gaagccacat 2040
actctcggtg cggctccggt ccctggatta cacccaggtg catggtcgac tacccgtata 2100
ggctttggca ctatccttgt accatcaatt acaccatatt caaagtcagg atgtacgtgg 2160
gaggggtcga gcacaggctg gaagcggcct gcaactggac gcggggcgaa cgctgtgatc 2220
tggaagacag ggacaggtcc gagctcagcc cgttgctgct gtccaccaca cagtggcagg 2280
tccttccgtg ttctttcacg accctgccag ccttgtccac cggcctcatc cacctccacc 2340
agaacattgt ggacgtgcag tacttgtacg gggtagggtc aagcatcgcg tcctgggcca 2400
ttaagtggga gtacgtcgtt ctcctgttcc ttctgcttgc agacgcgcgc gtctgctcct 2460
gcttgtggat gatgttactc atatcccaag cggaggcggc tttggagaac ctcgtaatac 2520
tcaatgcagc atccctggcc gggacgcacg gtcttgtgtc cttcctcgtg ttcttctgct 2580
ttgcgtggta tctgaagggt aggtgggtgc ccggagcggt ctacgccctc tacgggatgt 2640
ggcctctcct cctgctcctg ctggcgttgc ctcagcgggc atacgcactg gacacggagg 2700
tggccgcgtc gtgtggcggc gttgttcttg tcgggttaat ggcgctgact ctgtcgccat 2760
attacaagcg ctatatcagc tggtgcatgt ggtggcttca gtattttctg accagagtag 2820
aagcgcaact gcacgtgtgg gttccccccc tcaacgtccg gggggggcgc gatgccgtca 2880
tcttactcat gtgtgtagta cacccgaccc tggtatttga catcaccaaa ctactcctgg 2940
ccatcttcgg acccctttgg attcttcaag ccagtttgct taaagtcccc tacttcgtgc 3000
gcgttcaagg ccttctccgg atctgcgcgc tagcgcggaa gatagccgga ggtcattacg 3060
tgcaaatggc catcatcaag ttaggggcgc ttactggcac ctatgtgtat aaccatctca 3120
cccctcttcg agactgggcg cacaacggcc tgcgagatct ggccgtggct gtggaaccag 3180
tcgtcttctc ccgaatggag accaagctca tcacgtgggg ggcagatacc gccgcgtgcg 3240
gtgacatcat caacggcttg cccgtctctg cccgtagggg ccaggagata ctgcttgggc 3300
cagccgacgg aatggtctcc aaggggtgga ggttgctggc gcccatcacg gcgtacgccc 3360
agcagacgag aggcctccta gggtgtataa tcaccagcct gactggccgg gacaaaaacc 3420
aagtggaggg tgaggtccag atcgtgtcaa ctgctaccca aaccttcctg gcaacgtgca 3480
tcaatggggt atgctggact gtctaccacg gggccggaac gaggaccatc gcatcaccca 3540
agggtcctgt catccagatg tataccaatg tggaccaaga ccttgtgggc tggcccgctc 3600
ctcaaggttc ccgctcattg acaccctgta cctgcggctc ctcggacctt tacctggtca 3660
cgaggcacgc cgatgtcatt cccgtgcgcc ggcgaggtga tagcaggggt agcctgcttt 3720
cgccccggcc catttcctac ttgaaaggct cctcgggggg tccgctgttg tgccccgcgg 3780
gacacgccgt gggcctattc agggccgcgg tgtgcacccg tggagtggct aaagcggtgg 3840
actttatccc tgtggagaac ctagggacaa ccatgagatc cccggtgttc acggacaact 3900
cctctccacc agcagtgccc cagagcttcc aggtggccca cctgcatgct cccaccggca 3960
gcggtaagag caccaaggtc ccggctgcgt acgcagccca gggctacaag gtgttggtgc 4020
tcaacccctc tgttgctgca acgctgggct ttggtgctta catgtccaag gcccatgggg 4080

ttgatcctaa tatcaggacc ggggtgagaa caattaccac tggcagcccc atcacgtact 4140
ccacctacgg caagttcctt gccgacggcg ggtgctcagg aggtgcttat gacataataa 4200
tttgtgacga gtgccactcc acggatgcca catccatctt gggcatcggc actgtccttg 4260
accaagcaga gactgcgggg gcgagactgg ttgtgctcgc cactgctacc cctccgggct 4320
ccgtcactgt gtcccatcct aacatcgagg aggttgctct gtccaccacc ggagagatcc 4380
ccttttacgg caaggctatc cccctcgagg tgatcaaggg gggaagacat ctcatcttct 4440
gccactcaaa gaagaagtgc gacgagctcg ccgcgaagct ggtcgcattg ggcatcaatg 4500
ccgtggccta ctaccgcggt cttgacgtgt ctgtcatccc gaccagcggc gatgttgtcg 4560
tcgtgtcgac cgatgctctc atgactggct ttaccggcga cttcgactct gtgatagact 4620
gcaacacgtg tgtcactcag acagtcgatt tcagccttga ccctaccttt accattgaga 4680
caaccacgct cccccaggat gctgtctcca ggactcaacg ccggggcagg actggcaggg 4740
ggaagccagg catctataga tttgtggcac cgggggagcg cccctccggc atgttcgact 4800
cgtccgtcct ctgtgagtgc tatgacgcgg gctgtgcttg gtatgagctc acgcccgccg 4860
agactacagt taggctacga gcgtacatga acaccccggg gcttcccgtg tgccaggacc 4920
atcttgaatt ttgggagggc gtctttacgg gcctcactca tatagatgcc cacttttat 4980
cccagacaaa gcagagtggg gagaactttc cttacctggt agcgtaccaa gccaccgtgt 5040
gcgctagggc tcaagccccc cccccatcgt gggaccagat gtggaagtgt ttgatccgcc 5100
ttaaacccac cctccatggg ccaacacccc tgctatacag actgggcgct gttcagaatg 5160
aagtcaccct gacgcaccca atcaccaaat acatcatgac atgcatgtcg gccgacctgg 5220
aggtcgtcac gagcacctgg gtgctcgttg gcggcgtcct ggctgctctg gccgcgtatt 5280
gcctgtcaac aggctgcgtg gtcatagtgg gcaggatcgt cttgtccggg aagccggcaa 5340
ttatacctga cagggaggtt ctctaccagg agttcgatga gatggaagag tgctctcagc 5400
acttaccgta catcgagcaa gggatgatgc tcgctgagca gttcaagcag aaggccctcg 5460
gcctcctgca gaccgcgtcc cgccatgcag aggttatcac ccctgctgtc cagaccaact 5520
ggcagaaact cgaggtcttt tgggcgaagc acatgtggaa tttcatcagt gggatacaat 5580
acttggcggg cctgtcaacg ctgcctggta accccgccat tgcttcattg atggctttta 5640
cagctgccgt caccagccca ctaaccactg gccaaaccct cctcttcaac atattggggg 5700
ggtgggtggc tgcccagctc gccgcccccg gtgccgctac tgcctttgtg ggtgctggcc 5760
tagctggcgc cgccatcggc agcgttggac tggggaaggt cctcgtggac attcttgcag 5820
ggtatggcgc gggcgtggcg ggagctcttg tagcattcaa gatcatgagc ggtgaggtcc 5880
cctccacgga ggacctggtc aatctgctgc ccgccatcct ctcgcctgga gcccttgtag 5940
tcggtgtggt ctgcgcagca atactgcgcc ggcacgttgg cccgggcgag ggggcagtgc 6000
aatggatgaa ccggctaata gccttcgcct cccgggggaa ccatgtttcc cccacgcact 6060
acgtgccgga gagcgatgca gccgcccgcg tcactgccat actcagcagc ctcactgtaa 6120
cccagctcct gaggcgactg catcagtgga taagctcgga gtgtaccact ccatgctccg 6180
gttcctggct aagggacatc tgggactgga tatgcgaggt gctgagcgac tttaagacct 6240
ggctgaaagc caagctcatg ccacaactgc ctgggattcc ctttgtgtcc tgccagcgcg 6300
ggtatagggg ggtctggcga ggagacggca ttatgcacac tcgctgccac tgtggagctg 6360
agatcactgg acatgtcaaa aacgggacga tgaggatcgt cggtcctagg acctgcagga 6420
acatgtggag tgggacgttc cccattaacg cctacaccac gggcccctgt actccccttc 6480
ctgcgccgaa ctataagttc gcgctgtgga gggtgtctgc agaggaatac gtggagataa 6540
ggcgggtggg ggacttccac tacgtatcgg gtatgactac tgacaatctt aaatgcccgt 6600
gccagatccc atcgcccgaa tttttcacag aattggacgg ggtgcgccta cacaggtttg 6660
cgccccctg caagcccttg ctgcgggagg aggtatcatt cagagtagga ctccacgagt 6720
acccggtggg gtcgcaatta ccttgcgagc ccgaaccgga cgtagccgtg ttgacgtcca 6780
tgctcactga tccctcccat ataacagcag aggcggccgg gagaaggttg gcgagagggt 6840
cacccccttc tatggccagc tcctcggcta gccagctgtc cgctccatct ctcaaggcaa 6900
cttgcaccgc caaccatgac tcccctgacg ccgagctcat agaggctaac ctcctgtgga 6960

ggcaggagat gggcggcaac atcaccaggg ttgagtcaga gaacaaagtg gtgattctgg 7020
actccttcga tccgcttgtg gcagaggagg atgagcggga ggtctccgta cctgcagaaa 7080
ttctgcggaa gtctcggaga ttcgcccggg ccctgcccgt ctgggcgcgg ccggactaca 7140
acccccccgct agtagagacg tggaaaaagc ctgactacga accacctgtg gtccatggct 7200
gcccgctacc acctccacgg tcccctcctg tgcctccgcc tcggaaaaag cgtacggtgg 7260
tcctcaccga atcaacccta tctactgcct tggccgagct tgccaccaaa agttttggca 7320
gctcctcaac ttccggcatt acgggcgaca atacgacaac atcctctgag cccgcccctt 7380
ctggctgccc ccccgactcc gacgttgagt cctattcttc catgcccccc ctggaggggg 7440
agcctgggga tccggatctc agcgacgggt catggtcgac ggtcagtagt ggggccgaca 7500
cggaagatgt cgtgtgctgc tcaatgtctt attcctggac aggcgcactc gtcaccccgt 7560
gcgctgcgga agaacaaaaa ctgcccatca acgcactgag caactcgttg ctacgccatc 7620
acaatctggt gtattccacc acttcacgca gtgcttgcca aaggcagaag aaagtcacat 7680
ttgacagact gcaagttctg gacagccatt accaggacgt gctcaaggag gtcaaagcag 7740
cggcgtcaaa agtgaaggct aacttgctat ccgtagagga agcttgcagc ctgacgcccc 7800
cacattcagc caaatccaag tttggctatg gggcaaaaga cgtccgttgc catgccagaa 7860
aggccgtagc ccacatcaac tccgtgtgga aagacccttct ggaagacagt gtaacaccaa 7920
tagacactac catcatggcc aagaacgagg ttttctgcgt tcagcctgag aaggggggtc 7980
gtaagccagc tcgtctcatc gtgttccccg acctgggcgt gcgcgtgtgc gagaagatgg 8040
ccctgtacga cgtggttagc aagctccccc tggccgtgat gggaagctcc tacggattcc 8100
aatactcacc aggacagcgg gttgaattcc tcgtgcaagc gtggaagtcc aagaagaccc 8160
cgatggggtt ctcgtatgat acccgctgtt ttgactccac agtcactgag agcgacatcc 8220
gtacggagga ggcaatttac caatgttgtg acctggaccc ccaagcccgc gtggccatca 8280
agtccctcac tgagaggctt tatgttgggg gccctcttac caattcaagg ggggaaaact 8340
gcggctaccg caggtgccgc gcgagcggcg tactgacaac tagctgtggt aacaccctca 8400
cttgctacat caaggcccgg gcagcctgtc gagccgcagg gctccaggac tgcaccatgc 8460
tcgtgtgtgg cgacgactta gtcgttatct gtgaaagtgc gggggtccag gaggacgcgg 8520
cgagcctgag agccttcacg gaggctatga ccaggtactc cgccccccccc ggggaccccc 8580
cacaaccaga atacgacttg gagcttataa catcatgctc ctccaacgtg tcagtcgccc 8640
acgacggcgc tggaaagagg gtctactacc ttacccgtga ccctacaacc cccctcgcga 8700
gagccgcgtg ggagacagca agacacactc cagtcaattc ctggctaggc aacataatca 8760
tgtttgcccc cacactgtgg gcgaggatga tactgatgac ccatttcttt agcgtcctca 8820
tagccaggga tcagcttgaa caggctctta actgtgagat ctacggagcc tgctactcca 8880
tagaaccact ggatctacct ccaatcattc aaagactcca tggcctcagc gcattttcac 8940
tccacagtta ctctccaggt gaaatcaata gggtggccgc atgcctcaga aaacttgggg 9000
tcccgccctt gcgagcttgg agacaccggg cccggagcgt ccgcgctagg cttctgtcca 9060
gaggaggcag ggctgccata tgtggcaagt acctcttcaa ctgggcagta agaacaaagc 9120
tcaaactcac tccaatagcg gccgctggcc ggctggactt gtccggttgg ttcacggctg 9180
gctacagcgg gggagacatt tatcacagcg tgtctcatgc ccggccccgc tggttctggt 9240
tttgcctact cctgctcgct gcaggggtag gcatctacct cctccccaac cgatgaaggt 9300
tggggtaaac actccggcct cttaagccat ttcctgtttt ttttttttt ttttttttt 9360
ttttcttttt ttttttcttt cctttccttc tttttttcct ttctttttcc cttctttaat 9420
ggtggctcca tcttagccct agtcacggct agctgtgaaa ggtccgtgag ccgcatgact 9480
gcagagagtg ctgatactgg cctctctgca gatcatgt 9518

<210> 2
<211> 2984
<212> PRT

<213> Hepatitis C virus

<400> 2

```
Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
  1               5                  10                  15

Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
             20                  25                  30

Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala
         35                  40                  45

Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro
     50                  55                  60

Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly
 65                  70                  75                  80

Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp
                 85                  90                  95

Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
            100                 105                 110

Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys
        115                 120                 125

Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu
    130                 135                 140

Gly Gly Ala Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp
145                 150                 155                 160

Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile
                165                 170                 175

Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr
            180                 185                 190

Gln Val Arg Asn Ser Ser Gly Leu Tyr His Val Thr Asn Asp Cys Pro
        195                 200                 205

Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile Leu His Thr Pro
    210                 215                 220

Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser Arg Cys Trp Val
225                 230                 235                 240
```

```
Ala Val Thr Pro Thr Val Ala Thr Arg Asp Gly Lys Leu Pro Thr Thr
                245                 250                 255

Gln Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser Ala Thr Leu Cys
            260                 265                 270

Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Ser Val Phe Leu Val Gly
        275                 280                 285

Gln Leu Phe Thr Phe Ser Pro Arg Arg His Trp Thr Thr Gln Asp Cys
    290                 295                 300

Asn Cys Ser Ile Tyr Pro Gly His Ile Thr Gly His Arg Met Ala Trp
305                 310                 315                 320

Asp Met Met Met Asn Trp Ser Pro Thr Ala Ala Leu Val Val Ala Gln
                325                 330                 335

Leu Leu Arg Ile Pro Gln Ala Ile Met Asp Met Ile Ala Gly Ala His
            340                 345                 350

Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp
        355                 360                 365

Ala Lys Val Leu Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala Ile
    370                 375                 380

Gln Leu Ile Asn Thr Asn Gly Ser Trp His Ile Asn Ser Thr Ala Leu
385                 390                 395                 400

Asn Cys Asn Glu Ser Leu Asn Thr Gly Trp Leu Ala Gly Leu Phe Tyr
                405                 410                 415

Gln His Lys Phe Asn Ser Ser Gly Cys Pro Glu Arg Leu Ala Ser Cys
            420                 425                 430

Arg Arg Leu Thr Asp Phe Ala Gln Gly Trp Gly Pro Ile Ser Tyr Ala
        435                 440                 445

Asn Gly Ser Gly Leu Asp Glu Arg Pro Tyr Cys Trp His Tyr Pro Pro
    450                 455                 460

Arg Pro Cys Gly Ile Val Pro Ala Lys Ser Val Cys Gly Pro Val Tyr
465                 470                 475                 480

Cys Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg Ser Gly
                485                 490                 495
```

```
Ala Pro Thr Tyr Ser Trp Gly Ala Asn Asp Thr Asp Val Phe Val Leu
            500                 505                 510

Asn Asn Thr Arg Pro Pro Leu Gly Asn Trp Phe Gly Cys Thr Trp Met
        515                 520                 525

Asn Ser Thr Gly Phe Thr Lys Val Cys Gly Ala Pro Pro Cys Val Ile
    530                 535                 540

Gly Gly Val Gly Asn Asn Thr Leu Leu Cys Pro Thr Asp Cys Phe Arg
545                 550                 555                 560

Lys His Pro Glu Ala Thr Tyr Ser Arg Cys Gly Ser Gly Pro Trp Ile
                565                 570                 575

Thr Pro Arg Cys Met Val Asp Tyr Pro Tyr Arg Leu Trp His Tyr Pro
            580                 585                 590

Cys Thr Ile Asn Tyr Thr Ile Phe Lys Val Arg Met Tyr Val Gly Gly
        595                 600                 605

Val Glu His Arg Leu Glu Ala Ala Cys Asn Trp Thr Arg Gly Glu Arg
    610                 615                 620

Cys Asp Leu Glu Asp Arg Asp Arg Ser Glu Leu Ser Pro Leu Leu Leu
625                 630                 635                 640

Ser Thr Thr Gln Trp Gln Val Leu Pro Cys Ser Phe Thr Thr Leu Pro
                645                 650                 655

Ala Leu Ser Thr Gly Leu Ile His Leu His Gln Asn Ile Val Asp Val
            660                 665                 670

Gln Tyr Leu Tyr Gly Val Gly Ser Ser Ile Ala Ser Trp Ala Ile Lys
        675                 680                 685

Trp Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val
    690                 695                 700

Cys Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu Ala Ala
705                 710                 715                 720

Leu Glu Asn Leu Val Ile Leu Asn Ala Ala Ser Leu Ala Gly Thr His
                725                 730                 735

Gly Leu Val Ser Phe Leu Val Phe Phe Cys Phe Ala Trp Tyr Leu Lys
            740                 745                 750
```

```
Gly Arg Trp Val Pro Gly Ala Val Tyr Ala Leu Tyr Gly Met Trp Pro
        755                 760                 765

Leu Leu Leu Leu Leu Leu Ala Leu Pro Gln Arg Ala Tyr Ala Leu Asp
    770                 775                 780

Thr Glu Val Ala Ala Ser Cys Gly Gly Val Val Leu Val Gly Leu Met
785                 790                 795                 800

Ala Leu Thr Leu Ser Pro Tyr Tyr Lys Arg Tyr Ile Ser Trp Cys Met
                805                 810                 815

Trp Trp Leu Gln Tyr Phe Leu Thr Arg Val Glu Ala Gln Leu His Val
            820                 825                 830

Trp Val Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val Ile Leu
        835                 840                 845

Leu Met Cys Val Val His Pro Thr Leu Val Phe Asp Ile Thr Lys Leu
    850                 855                 860

Leu Leu Ala Ile Phe Gly Pro Leu Trp Ile Leu Gln Ala Ser Leu Leu
865                 870                 875                 880

Lys Val Pro Tyr Phe Val Arg Val Gln Gly Leu Leu Arg Ile Cys Ala
                885                 890                 895

Leu Ala Arg Lys Ile Ala Gly Gly His Tyr Val Gln Met Ala Ile Ile
            900                 905                 910

Lys Leu Gly Ala Leu Thr Gly Thr Tyr Val Tyr Asn His Leu Thr Pro
        915                 920                 925

Leu Arg Asp Trp Ala His Asn Gly Leu Arg Asp Leu Ala Val Ala Val
    930                 935                 940

Glu Pro Val Val Phe Ser Arg Met Glu Thr Lys Leu Ile Thr Trp Gly
945                 950                 955                 960

Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Asn Gly Leu Pro Val Ser
                965                 970                 975

Ala Arg Arg Gly Gln Glu Ile Leu Leu Gly Pro Ala Asp Gly Met Val
            980                 985                 990

Ser Lys Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ala Gln Gln
        995                1000                1005
```

Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp
1010 1015 1020

Lys Asn Gln Val Glu Gly Glu Val Gln Ile Val Ser Thr Ala Thr Gln
1025 1030 1035 1040

Thr Phe Leu Ala Thr Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His
1045 1050 1055

Gly Ala Gly Thr Arg Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln
1060 1065 1070

Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln
1075 1080 1085

Gly Ser Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr
1090 1095 1100

Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp
1105 1110 1115 1120

Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly
1125 1130 1135

Ser Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu
1140 1145 1150

Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe
1155 1160 1165

Ile Pro Val Glu Asn Leu Gly Thr Thr Met Arg Ser Pro Val Phe Thr
1170 1175 1180

Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His
1185 1190 1195 1200

Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala
1205 1210 1215

Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala
1220 1225 1230

Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly Val Asp
1235 1240 1245

Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ser Pro Ile
1250 1255 1260

```
Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly
1265                1270                1275                1280

Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ala
                1285                1290                1295

Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala
            1300                1305                1310

Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val
        1315                1320                1325

Thr Val Ser His Pro Asn Ile Glu Glu Val Ala Leu Ser Thr Thr Gly
    1330                1335                1340

Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly
1345                1350                1355                1360

Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu
                1365                1370                1375

Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg
            1380                1385                1390

Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Val Val Val
        1395                1400                1405

Ser Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val
    1410                1415                1420

Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp
1425                1430                1435                1440

Pro Thr Phe Thr Ile Glu Thr Thr Thr Leu Pro Gln Asp Ala Val Ser
                1445                1450                1455

Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr
            1460                1465                1470

Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser
        1475                1480                1485

Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr
    1490                1495                1500

Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly
1505                1510                1515                1520
```

Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly Val Phe Thr
1525 1530 1535

Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ser
1540 1545 1550

Gly Glu Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala
1555 1560 1565

Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu
1570 1575 1580

Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg
1585 1590 1595 1600

Leu Gly Ala Val Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys
1605 1610 1615

Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr
1620 1625 1630

Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu
1635 1640 1645

Ser Thr Gly Cys Val Val Ile Val Gly Arg Ile Val Leu Ser Gly Lys
1650 1655 1660

Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Gln Glu Phe Asp Glu
1665 1670 1675 1680

Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln Gly Met Met
1685 1690 1695

Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala
1700 1705 1710

Ser Arg His Ala Glu Val Ile Thr Pro Ala Val Gln Thr Asn Trp Gln
1715 1720 1725

Lys Leu Glu Val Phe Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly
1730 1735 1740

Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile
1745 1750 1755 1760

Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro Leu Thr Thr
1765 1770 1775

Gly Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln
1780 1785 1790

Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala
1795 1800 1805

Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu Val Asp Ile
1810 1815 1820

Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys
1825 1830 1835 1840

Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu
1845 1850 1855

Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val Cys Ala
1860 1865 1870

Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp
1875 1880 1885

Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro
1890 1895 1900

Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val Thr Ala Ile
1905 1910 1915 1920

Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu His Gln Trp
1925 1930 1935

Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp
1940 1945 1950

Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys Thr Trp Leu
1955 1960 1965

Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe Val Ser Cys
1970 1975 1980

Gln Arg Gly Tyr Arg Gly Val Trp Arg Gly Asp Gly Ile Met His Thr
1985 1990 1995 2000

Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys Asn Gly Thr
2005 2010 2015

Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp Ser Gly Thr
2020 2025 2030

Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro Leu Pro Ala
2035 2040 2045

Pro Asn Tyr Lys Phe Ala Leu Trp Arg Val Ser Ala Glu Glu Tyr Val
2050 2055 2060

Glu Ile Arg Arg Val Gly Asp Phe His Tyr Val Ser Gly Met Thr Thr
2065 2070 2075 2080

Asp Asn Leu Lys Cys Pro Cys Gln Ile Pro Ser Pro Glu Phe Phe Thr
2085 2090 2095

Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro Cys Lys Pro
2100 2105 2110

Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His Glu Tyr Pro
2115 2120 2125

Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val Ala Val Leu
2130 2135 2140

Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Ala Ala Gly
2145 2150 2155 2160

Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Met Ala Ser Ser Ser Ala
2165 2170 2175

Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr Ala Asn His
2180 2185 2190

Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln
2195 2200 2205

Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn Lys Val Val
2210 2215 2220

Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp Glu Arg Glu
2225 2230 2235 2240

Val Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg Phe Ala Arg
2245 2250 2255

Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu Val Glu
2260 2265 2270

Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His Gly Cys Pro
2275 2280 2285

```
Leu Pro Pro Pro Arg Ser Pro Pro Val Pro Pro Pro Arg Lys Lys Arg
   2290                2295                2300

Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu Ala Glu Leu
2305                2310                2315                2320

Ala Thr Lys Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile Thr Gly Asp
                2325                2330                2335

Asn Thr Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys Pro Pro Asp
            2340                2345                2350

Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly Glu Pro
        2355                2360                2365

Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser Ser Gly
   2370                2375                2380

Ala Asp Thr Glu Asp Val Val Cys Cys Ser Met Ser Tyr Ser Trp Thr
2385                2390                2395                2400

Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys Leu Pro Ile
                2405                2410                2415

Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr Ser
            2420                2425                2430

Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val Thr Phe Asp
        2435                2440                2445

Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu Lys Glu Val
   2450                2455                2460

Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser Val Glu Glu
2465                2470                2475                2480

Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys Phe Gly Tyr
                2485                2490                2495

Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val Ala His Ile
            2500                2505                2510

Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Ser Val Thr Pro Ile Asp
        2515                2520                2525

Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu Lys
   2530                2535                2540
```

```
Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly Val
2545                2550                2555                2560

Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Lys Leu Pro
                2565                2570                2575

Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln
            2580                2585                2590

Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys Thr Pro Met
        2595                2600                2605

Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu Ser
    2610                2615                2620

Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp Leu Asp Pro
2625                2630                2635                2640

Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu Tyr Val Gly
                2645                2650                2655

Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr Arg Arg Cys
            2660                2665                2670

Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys
        2675                2680                2685

Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu Gln Asp Cys
    2690                2695                2700

Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser Ala
2705                2710                2715                2720

Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr Glu Ala Met
                2725                2730                2735

Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro Glu Tyr Asp
            2740                2745                2750

Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His Asp
        2755                2760                2765

Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro
    2770                2775                2780

Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn Ser
2785                2790                2795                2800
```

```
Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp Ala Arg Met
                2805                2810                2815

Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg Asp Gln Leu
            2820                2825                2830

Glu Gln Ala Leu Asn Cys Glu Ile Tyr Gly Ala Cys Tyr Ser Ile Glu
        2835                2840                2845

Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly Leu Ser Ala
    2850                2855                2860

Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala Ala
2865                2870                2875                2880

Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp Arg His Arg
                2885                2890                2895

Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Arg Gly Gly Arg Ala Ala
            2900                2905                2910

Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu Lys
        2915                2920                2925

Leu Thr Pro Ile Ala Ala Ala Gly Arg Leu Asp Leu Ser Gly Trp Phe
    2930                2935                2940

Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val Ser His Ala
2945                2950                2955                2960

Arg Pro Arg Trp Phe Trp Phe Cys Leu Leu Leu Leu Ala Ala Gly Val
                2965                2970                2975

Gly Ile Tyr Leu Leu Pro Asn Arg
            2980
```

<210> 3
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 3
acgcgtctgc agcttaatgg cccaggacgc gatgcttg 38

<210> 4
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 4
acgcgtagat cttaccaagt gcgcaattcc tcgggg 36

<210> 5
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 5
tcagttggat agcgtcgacg ccggcaaata gcag 34

<210> 6
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 6
cgtcgacgcg atccaactga tcaacaccaa cggc 34

<210> 7
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

Primer

<400> 7
cttgtaccat caattacacc atattc 26

<210> 8
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 8
gatagtgcca atgcctatac ggg 23

<210> 9
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 9
cgtataggca ttggcactat ccttgtacca tcaattacac c 41

<210> 10
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 10
ggtgtaattg atggtacaag gatagtgcca atgcctatac g 41

<210> 11
<211> 41
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 11
tttttttgc ggccgcatgg tggggaactg ggcgaaggtc c 41

<210> 12
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 12
acgcgtaagc ttctattact cggacctgtc cctgtcttcc ag 42

INTERNATIONAL SEARCH REPORT

Interna al Application No
PCT/US 00/25987

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/40 C12N5/10 C12N7/01 C07K14/18 C07K16/10
A61K31/70 A61K39/29

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
WPI Data, BIOSIS, EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	FORNS X. ET AL.: "Characterization of modified Hepatitis C virus E2 protein expressed on the cell surface." VIROLOGY, vol. 274, 15 August 2000 (2000-08-15), pages 75-85, XP002156408 the whole document ---	1-60
X	WO 96 40764 A (US HEALTH) 19 December 1996 (1996-12-19) abstract page 10, line 27 - line 32 --- -/--	1,10,12, 14,16, 18,19, 21,23, 25-27, 29,31, 32,34, 35,37-39

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 January 2001	24/01/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Galli, I

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Interna al Application No
PCT/US 00/25987

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 99 04008 A (US HEALTH) 28 January 1999 (1999-01-28) abstract sequence of E2 from isolate H77C	1-60
A	WO 94 26306 A (CHIRON CORP) 24 November 1994 (1994-11-24) abstract	1-60
A	MALET I ET AL: "YELLOW FEVER 5' NONCODING REGION AS A POTENTIAL ELEMENT TO IMPROVE HEPATITIS C VIRUS PRODUCTION THROUGH MODIFICATION OF TRANSLATIONAL CONTROL" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS,ACADEMIC PRESS INC. ORLANDO, FL,US, vol. 253, no. 2, 18 December 1998 (1998-12-18), pages 257-264, XP000961230 ISSN: 0006-291X the whole document	3-9
A	YI M. ET AL.: "Delineation of regions important for heteromeric association of Hepatitis C Virus E1 and E2." VIROLOGY, vol. 231, 1997, pages 119-129, XP002156409 the whole document	40-60

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern: al Application No
PCT/US 00/25987

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9640764 A	19-12-1996	US 6110465 A	29-08-2000
		AU 718503 B	13-04-2000
		AU 6157996 A	30-12-1996
		CA 2221313 A	19-12-1996
		EP 0832114 A	01-04-1998
WO 9904008 A	28-01-1999	US 6153421 A	28-11-2000
		AU 8488998 A	10-02-1999
		EP 0998567 A	10-05-2000
WO 9426306 A	24-11-1994	CA 2162557 A	24-11-1994
		EP 0697888 A	28-02-1996
		JP 8510240 T	29-10-1996